REPORT RESUMES

ED 010 375 24

EFFECTS OF LIKED AND DISLIKED TEACHERS ON STUDENT BEHAVIOR.

BY- CARPENTER, FINLEY HADDAN, EUGENE E.
UNIVERSITY OF MICHIGAN, ANN ARBOR

REPORT NUMBER CRP-2450 PUB DATE 66
REPORT NUMBER BR-5-0335
EDRS PRICE MF-$0.18 HC-$3.80 95P.

DESCRIPTORS- *TEACHER CHARACTERISTICS, *STUDENT BEHAVIOR, LEARNING ACTIVITIES, COMPARATIVE ANALYSIS, *STUDENT OPINION, LEARNING MOTIVATION, SENSORY EXPERIENCE, TEACHER EVALUATION, *STUDENT TEACHER RELATIONSHIP, *PERSONALITY STUDIES, ANN ARBOR, MICHIGAN

RESEARCH WAS CONDUCTED TO CONTRAST THE EFFECTS OF TWO TYPES OF TEACHERS, LIKED AND DISLIKED, ON THE LEARNING BEHAVIOR OF THEIR STUDENTS. TEACHERS PRESENTED MESSAGES BY FILM, BY TAPE, AND IN PERSON IN EXPERIMENTAL CLASSROOMS TO STUDENTS FITTED WITH FINGER ELECTRODES. CHANGES IN ELECTRICAL RESISTANCE WERE RECORDED OF GALVANIC SKIN RESPONSES. ACHIEVEMENT TESTS WERE ALSO ADMINISTERED. MEASUREMENTS INCLUDED (1) PHYSIOLOGICAL AROUSAL, (2) RATINGS OF THE TEACHER, (3) RATINGS OF THE SUBJECT MATTER, (4) SCORES ON ACHIEVEMENT TESTS, AND (5) SCORES ON TESTS OF INFERENCE. RESPONSES, RATINGS, AND SCORES OF BOTH COLLEGE AND HIGH SCHOOL STUDENTS WERE STUDIED. AMONG THE SEVERAL FINDINGS WERE (1) STUDENTS RECEIVED SIGNIFICANTLY HIGHER SCORES ON TESTS BOTH OF FACTS AND OF INFERENCE UNDER THE "DISLIKED" TEACHERS WHEN SUBJECT MATTER WAS PRESENTED IN PERSON AND ON FILM, AND (2) STUDENTS ACHIEVED HIGHER SCORES UNDER "LIKED" TEACHERS ONLY WHEN SUBJECT MATTER WAS PRESENTED BY TAPE RECORDING. FOR ANY GIVEN PRESENTATION, NO RELATIONSHIP WAS FOUND BETWEEN HOW STUDENTS RATED THE TEACHER AND AMOUNT LEARNED. SUBJECT MATTER RATINGS APPEARED TO HAVE MUCH MORE BEARING ON ACHIEVEMENT THAN DID THE FACTOR OF TEACHER LIKEABILITY. (RS)

ED010375

5-033

EFFECTS OF LIKED AND DISLIKED TEACHERS ON STUDENT BEHAVIOR

Cooperative Research Project No. 2450

Finley Carpenter
Eugene E. Haddan

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

The University of Michigan

EFFECTS OF LIKED AND DISLIKED TEACHERS ON STUDENT BEHAVIOR

Cooperative Research Project No. 2450

Finley Carpenter
Project Director and Principal Investigator

Eugene E. Haddan
Research Associate

The University of Michigan
Ann Arbor, Michigan
1966

The research reported here was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

PRECEDING PAGE BLANK-NOT FILMED

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF PLATES

ABSTRACT

PREFACE

CHAPTER

1. INTRODUCTION AND RATIONALE 1
    - Teacher Attitudes 6
    - Teacher Interests 7
    - Teacher-Pupil Relations 7
    - Cognitive Clearness and Affective Tone 9
    - Definition of Terms 11
    - The Hypotheses 13
    - Limitations of the Study 15
    - Assumptions 16

2. CONSIDERATIONS ON INSTRUMENTATION 17
    - The Semantic Differential 17
    - The Galvanic Skin Response 18
        - Reliability 20
        - Equipment types 20
        - Electrodes 22
        - Electrode paste 23
        - Artifacts 23
        - Other sources of error 24
        - Units of measurement 24

3. THE PILOT EXPERIMENT 26
    - Design of the Study 26
    - Measurements 37
    - Results 38
    - Achievement Under the Various Treatments 41
    - Summary 44

4. THE SECOND EXPERIMENT 46
    - Review of the Aims of the Study 46
    - Changes Made as a Result of the Pilot Study 46
        - Changes in the electronic equipment 46

TABLE OF CONTENTS (Concluded)

CHAPTER Page

Assessment of achievement tests 47
Changes in the sample of students 47
Selection of teachers 48
Summary of the Procedures 48
Results in Relation to the Hypotheses 50
Summary of results in relation to the predictions 58
Further Results 58
Discussion 61

5. SUMMARY AND CONCLUSIONS 66
Summary 66
Conclusions 69
Recommendations 70

APPENDIX A. SEMANTIC DIFFERENTIAL SCALES 72

APPENDIX B. INSTRUCTIONS READ TO EXPERIMENTAL SUBJECTS 73

APPENDIX C. DIRECTIONS FOR TAKING THE MULTIPLE-CHOICE TESTS 74

APPENDIX D. IDENTIFICATION OF SUBJECT MATTER FOR THE SECOND EXPERIMENT 75

APPENDIX E. A SMALL SEGMENT OF A TYPICAL GSR RECORDING ON A STRIP CHART 76

REFERENCES 77

LIST OF TABLES

Table Page

1. Assignment of Messages (Units) to Modes and Polarities 36

2. Measurements 37

3. Obtained Mean Scores of GSR Frequencies Under the Treatments 38

4. Analysis of Mean Differences Between GSR Scores Across Modes of Presentation (t-ratios) 39

5. Mean Ratings of Teachers Across Modes on Semantic Differential Scales 39

6. Mean Ratings of Subject Matter Under Positive and Negative Presentations 40

7. Means of Achievement Scores Under the Negative Teacher by Mode of Presentation 41

8. Analysis of Significance between Means of Achievement Scores by Mode of Presentation Under the Negative Teacher 42

9. Means ofAchievement Under the Positive Teacher by Mode of Presentation 42

10. Analysis of Significance between Means of Achievement Scores by Mode of Presentation Under the Positive Teacher (t-ratios) 43

11. Analysis of Significance of Differences between Achievement Scores Under the Positive Teacher vs. the Negative Teacher 43

12. Reliability Coefficients of Achievement Tests Computed by the Kuder-Richardson Technique 47

13. Obtained Means of GSR Frequencies Under the Various Treatments 57

14. Mean Ratings of Teachers Across Modes on Semantic Differential Scales 52

15. Means and Standard Deviations of Achievement Scores Under Negative Conditions by Modes of Presentation 53

LIST OF TABLES (Concluded)

Table | Page

16. Means and Standard Deviations of Achievement Scores Under Positive Conditions by Mode of Presentation 54

17. Mean Differences Between Factual Test Scores, Showing t-Ratios and Significance Levels 55

18. Mean Differences Between Inference Test Scores, with the t-Ratio Values and Significance Levels 56

19. Amount of Physiological Arousal (GSRs) and Learning Stimulation (Live and Film)—Scores Shown Separately for Each Teacher 60

20. Amount of Physiological Arousal (GSRs) and Learning Under Presentations Limited to Audio Stimulation (Tape)—Scores Shown Separately for Each Teacher 60

LIST OF PLATES

Plate Page

I. Bausch and Lomb Servo Recorders for Making GSR Tracings 14

II. The Experimental Classroom 27

III. Student Desk and Electrodes 28

IV. Electronic Mixer, Amplifier and Recorders 29

V. Motion-picture Screen and Loudspeaker 30

VI. Device for Recording Affective Responses of Students During Learning Sessions 33

PRECEDING PAGE BLANK NOT FILMED

ABSTRACT

The aim of this study was to contrast the effects of two different kinds of teacher on the learning behavior of students. One kind of teacher was rated high on such scales as "likeable-annoying," "good-bad," "friendly-unfriendly." The other kind of teacher received consistently low ratings on the same scales. Six scales were used to rate the teachers and another six scales were used to rate the subject matter.

Each teacher presented three messages in an experimental classroom accommodating four students at a time. Students were fitted with pairs of finger electrodes so that changes in electrical resistance of the skin could be amplified and recorded in an adjoining room. The purpose was to obtain a record of galvanic skin responses (GSRs) during the course of learning, and to relate the GSR record of each student to his ratings of the teacher and of the subject matter. Achievement tests were made up that measured both the amount of factual learning and the ability to respond to items that required logical inference. Measurements of dependent variables therefore included: physiological arousal as indicated by the GSRs, ratings of the teacher, ratings of the subject matter, scores on factual achievement tests, and scores on tests of inference.

Messages were presented by means of tape recording, film, and in person, or live, delivery. Each teacher delivered three messages: one by tape, one by film, and one in person.

We examined the responses of students to the positive and negative teachers under each mode of delivery, and compared patterns of physiological arousal with achievement and with ratings of both the subject matter and the teacher. Achievement scores were also studied in comparison with ratings given to the teachers.

Two experiments were conducted: the first was a pilot study to determine what changes were required to improve data collection, and the second was run on the basis of the improvements discovered in the first.

In the pilot study, a sample of college students was used. They were mostly sophomores enrolled in an introductory psychology course. In the second experiment, 21 high school students were studied.

Messages were all about equally balanced in the loading of emotional words and all of approximately the same length. The order of presentation by instructor and by means of delivery was randomized for each group of four students to minimize the possible effect of a fixed sequence.

Data were used to test the following hypotheses:

(a) Messages presented in person will produce more arousal than tape or film presentations. The taped delivery will produce the least arousal because the fewest extraneous stimuli will bear upon the student.

(b) The amount of physiological arousal will be influenced by the mode of delivery. The greatest GSR response will occur when messages are delivered in person, and the least when presented by tape recording. (Although we did not specifically hypothesize that greater arousal would be produced by the disliked teacher than the liked teacher, we expected that result to occur.)

(c) Under the negative teacher, achievement scores will be highest by taped delivery and lowest by the live presentation. Dislike for the negative teacher will be greatest when he delivers the message in person, and the high level of dislike will produce "emotional noise" in the student that will interfere with his achievement.

(d) Under the positive teacher, a similar gradient of achievement will occur as under the negative teacher, but its slope will be less pronounced. Even strong liking for the teacher will produce "emotional noise" that will interfere with achievement, especially as measured by inference tests, because the strong positive appeal will produce uncritical acceptance of the subject matter. Consequently, the mental set of uncritical acceptance will tend to lower scores on tests requiring critical analysis.

(e) Ratings of teachers and subject matter will change according to the mode of presentation. The positive teacher will be rated highest when he presents material in person, and lowest when he presents it by tape recording. The order of ratings will be reversed for the negative teacher; that is, lowest ratings will be given for the live delivery, and highest ratings for tape.

Results of the study did not confirm our predictions. Further examination of the data, however, appeared to justify our efforts. The additional findings, which seem significant in a practical sense, are summarized below:

(a) Students received significantly higher scores on tests both of facts and of inference under the negative teacher than under the positive teacher when subject matter was presented in person and on film.

(b) Students achieved higher scores under the positive teacher than under the negative teacher only when material was delivered by tape recording.

(c) The negative teacher produced significantly higher physiological arousal than the positive teacher.

(d) Slow learners, who were so named because of their consistently low level of educational development according to their school records, produced erratic patterns of physiological arousal. In short, their patterns of arousal were less consistent than those of rapid learners.

(e) For any given presentation, there was no relationship between how students rated the teacher and amount learned.

(f) The most promising index of teaching effectiveness was the difference score between ratings of the teacher and ratings of the subject matter. Under the liked teacher, achievement was highest for students who showed little difference between teacher and content ratings. Under the disliked teacher, the greater the disparity in favor of subject matter, the higher the achievement.

Our conclusions are necessarily limited in generality because of the nature of our sample. We believe, however, the following conclusions are reasonably valid.

(a) When a teacher is strongly liked by students, it does not necessarily follow that the subject matter will be equally liked.

(b) When students rate subject matter very low, their level of achievement also tends to be low.

(c) The teacher who fails to arouse the feelings of students during instruction is less likely to stimulate high achievement than the teacher who creates considerable arousal. That appears to hold for both liked and disliked teachers.

(d) When content is delivered by tape recording, it is important that the speaker's voice be pleasant and that his delivery be relatively free from hesitation.

(e) Students can endure considerable negative stimulation from a teacher and still learn quite well when the teacher uses ample visual cues to strengthen his vocal message.

We think that our study has certain practical implications that can be tested by further research. The following recommendations are suggested as hypotheses for those interested in conducting such studies.

(a) When students have a low opinion of the courses offered by a department, as compared with other courses in the university, to restructure the disliked courses until student opinion is raised significantly. If bright students, in particular, fail to respect the intellectual value of the subject matter, they will give it minimal attention and concentrate their efforts on other courses which are challenging. Course modification may be more important than the hiring of popular teachers.

(b) When new methods of teaching are introduced for comparison, the results are likely to be misleading when the teaching innovation requires considerable readjustment on the part of students. A fair comparison can be made only after students have been given sufficient time to adapt to the change. Therefore, when methods of instruction involve radical departures, it seems particularly important to include a pilot study to determine the time required for adaptation.

(c) For those who are interested in measuring teaching effectiveness, we recommend that semantic differential scales be used to rate both teachers and subject matter. Teaching effectiveness is perhaps more adequately measured by using the difference score between ratings of the teacher and subject matter than by ignoring that derived score.

(d) Descriptive information of the teaching-learning process adequate for the development of a useful theory of teaching apparently requires the establishment of sets of three-fold relationships, including measures of the following classes of variables:

Input variables that include aspects of the teacher, content, and mode of presentation;

Process variables that include physiological arousal, a measure of feeling tone during the learning session, and cognitive reactions of students at the time of instruction (special computer-based research components offer the most promise for collecting data during learning sessions);

Output variables, including scores on tests of different levels of learning (knowledge, analysis, application, evaluation) and affective responses toward the content and teacher.

PREFACE

Educational development of the student is no doubt the result of many influences. It has long been held that teachers stimulate in students feeling tones that have much to do with attitudes, aspirations, interests, and intellectual achievement. Quantitative relations, however, have not yet been established to form a stable picture of the interaction of forces between those that are largely emotional and those that are mainly informative. The experiments reported in this volume were attempts to move nearer to specifying the kind and degree of relationships that exist among such variables as: student ratings of the teacher, ratings of the subject matter, physiological arousal during learning, media for presenting subject matter, and achievement.

The experimental approach to the study of classroom processes contains both advantages and shortcomings. The weaknesses are found in the limitations of control and measurement required for getting the kind of data suitable for describing the central relationships. Problems are further multiplied by the great variety and range of individual differences, which reflect the many kinds of behavior that result from any given educational stimulus. It is believed, however, that such problems will recede in the face of technological progress in the control and measurement of the teaching-learning complex. In the meantime, it is believed that current methods of experimentation can help to narrow the scope of likely factors that influence the feelings and achievements of students. It is within this perspective that the present study was conceived.

Special effort was made to write the report so that a relatively large number of readers could follow it without becoming hopelessly lost in technical jargon. We sacrificed some rigor and precision for what we hope is reasonably clear communication to teachers, school administrators, and interested laymen who have a limited background in the technicalities of educational research. We believe that the nature of our study justified this effort.

The success of most if not all research projects depends upon the contributions of a number of people. We want to recognize those individuals whose assistance and cooperation were central to the completion of the study: Ford Lemler and Aubert Lavastida, both of the Audio-Visual Education Center, The University of Michigan, for their competent counsel and direction of the film productions; Curtis Coleman, an electronics engineer, for his expert service in the design and construction of the electronic components used in measuring physiological arousal; Daniel Lirones, Manager of Film Central, for his production of the photographs used in the report; and Leslie Schwab, a budding research genius, for his patient, persistent, and fruitful work in locating valuable findings beyond those required to test the original hypotheses. The reader will later note that the main value of the study lies in the findings beyond those used to assess the preconceived predictions.

CHAPTER ONE: INTRODUCTION AND RATIONALE

There is much about human learning that remains a mystery. The bewildering interplay of biological and mental forces that somehow bring about the miracle of human development is a dark continent in the psychological world. Almost since the origin of learning theory in Aristotle's day, many conflicting guesses have been voiced about how the human being develops from a tiny cell to the most complicated of all organisms.

Among the influences that contribute to the educational development of a child are conditions arousing emotional reactions that result in the establishment of attitudes; such attitudes predispose the child to behave in set ways toward many aspects of his environment. Other influences of education include promotion of understanding and application of the knowledge contained in any discipline. Learning subject matter can be either enhanced or retarded by the attitudes that have been formed through earlier experiences.

Educational psychologists differ somewhat in the emphasis they place upon the role of emotion in the educational process. None deny its influence; but some believe that it is so fundamental that unless teachers have a good grasp of the emotional development of the child they cannot perform their duties effectively. Other psychologists, however, put the stress on those relations that are directly involved in the acquisition of knowledge and in other cognitive processes. They study such things as the value of practice in learning skills, meaningfulness of subject matter, techniques of cueing, feedback, and the structural variables of verbal communication. In educational psychology the pendulum swings between emphasis upon the emotional and cognitive aspects of human learning. Recently the pendulum has swung toward the cognitive side because of research on teaching machines, programming, game theory, and problem solving. Signs of compromise between the two points of emphasis are seen in the growing amount of research in which both cognition and emotion are involved. It now seems apparent that both factors must be studied simultaneously if we are to make headway in describing the intricate processes of teaching and learning as they occur in education.

In the classroom there are many influences that play upon the student. Psychologically it is convenient to think of those influences as different classes of stimulation. It is reasonable to divide classroom stimuli into two large sections: those that tend to arouse emotional reactions and those that contain information and that bring about no significant change in the way the student feels. The straight-forward fact, for example, that Columbus is the capital of Ohio should not be expected to either depress or delight students to any great extent. Yet, the mannerisms of the teacher, his enthusiastic or monotonous way of giving that information may provide emotional side effects that bear upon the student's perception and upon how long he retains the information. The interaction between stimuli that arouse feelings and those that provide factual content (produce cognition) is likely to have different effects upon students of different personality types. We still have not discovered the best way of sorting students so that they can be put under the kind of teaching and learning conditions that fit their modes of perception, thinking, and feeling. At present we make only primitive groupings, based largely upon lack of knowledge, in an effort to get the most mileage from our teaching efforts.

The study reported here grew out of results of earlier research by the director of the study. In that research, a simple interview technique was used under informal conditions designed to encourage the student to talk about his opinions of teachers and courses, and about his attitude towards school. The upshot of the study centered on differences between successful and unsuccessful students. Those who were failing, or on the borderline of failure, spoke emotionally of teachers in mostly negative terms. These students had similar feelings about the courses that were taught by teachers whom they disliked. Successful students, on the other hand, often expressed different feelings towards the teacher and towards the course taught by that teacher. While many successful students had strong negative feelings about some teachers, they frequently voiced positive feelings about the courses taught by the disliked teachers.

The obvious interpretation of the results was that unsuccessful students were probably victims of over-extended feelings from the teacher to the subject matter. Technically, we may call that probable phenomenon "overgeneralization of negative affect," meaning that because the student disliked the teacher he also learned to dislike the course. It may be that many unsuccessful students are victims of such overgeneralization. Successful students apparently had the ability to discriminate between the teacher and the course and could maintain a positive feeling toward the course even when they disliked the teacher. The study reported here was designed to test predictions drawn partly from the above speculations.

Teachers perform many functions, most of which are not clearly understood when it come to judging their effects upon individual students. We still do not have a good scientific catalog of the things that teachers do in the process of instruction. If we did, we could improve our progress in building a useful theory of teaching, a much needed tool for increasing the amount of pay-off per unit of investment in education.

It is convenient and worthwhile to regard all of the many functions involved in teaching under the single term, stimulation. The term is useful because it is impossible to conceive of any influence that a teacher has upon a student that does not involve some form of stimulation. By simply classifying all teaching efforts under stimulation, we do not, however, solve any problems; we only introduce a concept for making a plausible approach to the problems as a first step toward clarifying certain issues that otherwise might remain too ambiguous for fruitful research.

Exploring the stimulus concept, we find in the teacher's effort to give students new information two kinds of stimulus: a kind that embodies only the information that the teacher intends to convey to students, and a second sort that is peripheral or extraneous to the information that the teacher wants students to acquire. We can accordingly call them essential and extraneous stimuli. For example, the teacher could point to a map of Central America and say, "Here is the Isthmus of Panama, a narrow strip of land that connects North and South America." The essential stimulus can be regarded as the quoted sentence plus the visual configuration on the map comprising the isthmus. Teaching, however, is never confined only to those stimuli that constitute the subject matter. It involves some context, some other stimuli external to those that we have named as essential. As noted before, the mannerisms of the teacher, his attitude towards the class, his attitude towards the subject matter, his appearance, and the traits of his personality all constitute stimuli that can influence the student.

In general, the problem described in this report was so developed and designed that an estimate could be made of the effects of extraneous stimuli on the learning of content presented by teachers who were liked by students as opposed to those who were disliked. The problem is more complicated than this suggests, and its other facets will be described in detail in the chapter on procedure.

Human beings have much in common in their early care and training. Typically, the young infant is cared for by his mother, who provides for his needs and tries to encourage those forms of behavior that will give him a good start in adjusting to family life and in meeting the problems imposed by society when his experiences extend beyond the home. There are many concepts in psychology that are used to describe the early development of the child. In this study, we have selected a minimum of technical terms in the interest of clear communication. We believe that it is sufficient, at least to the aims of this research, to say that the main influences on human behavior are of two kinds: those forces in the environment that are rewarding to the person and those that are punishing. Although rewards and punishments may not cover all the significant influences on human behavior, we believe that the many experiences that fall into these two classes actually serve as significant factors in shaping the person's attitudes, his aspirations, his beliefs, interests, feelings, perceptions, and other emotional and cognitive aspects of his personality. If we can accept the foregoing statement as a basic assumption about the psychology of human behavior, we can develop a meaningful and consistent viewpoint that can promise to lead us into valuable areas of investigation, particularly in the fields of education.

As said before, the mother is usually the central figure in the early training of an infant. And in carrying out the tasks in such training, the

the mother uses many forms of reward and later introduces certain warnings and punishments to teach the child the do's and don't's of acceptable behavior. Both reward and punishment become more complex as the child grows, and he learns that vocal communication carries many signals of both reward and punishment. Consequently, we can say that much of his behavior gradually comes under the influence of verbal expressions from his social environment.

Although the term "reward" is meaningful in ordinary discourse, it is somewhat too subjective to convey our meaning. Consequently, we prefer the concept "reinforcement," a technical term that is roughly equivalent to "reward," but that has certain advantages over the latter. In the psychology of behavior analysis, reinforcement is used to refer to environmental influences that strengthen particular kinds of behavior. For example, the child who recites correctly in response to the teacher's request is reinforced when the teacher acknowledges the accuracy of the response and commends the student on the quality of his performance. A hungry animal can be taught to do a variety of tricks through proper cueing that indicates to the animal which performance leads to food. Dogs, monkeys, horses, porpoises, and many other animals can be trained to respond to signals that terminate in reinforcement. To say that the animal is rewarded suggests that we can perceive its inner feelings of satisfaction, which can actually only be inferred, but to say that the animal is "reinforced," as evidenced by his increased tendency to perform the same act under similar conditions, does not involve the difficulty introduced by the term "reward." A catalog can be compiled of reinforcers (those things and processes that produce reinforcement) that occur in school, such as good grades, gold stars, awards, commendations, and recognition of merit by fellow students.

The concept "reinforcement" has proved to be a useful term for describing and accounting for learning both on the animal and human levels. The significance of the term in this study springs from the notion that it offers a reasonable basis for speaking about the origin of positive feelings in teaching-learning situations. For example, if a teacher is perceived as pleasing and frequently recognizes the efforts and achievements of the student, it is likely that the student will form a positive feeling towards that teacher. And in some students the good opinion of the teacher may lead to an accepting attitude towards the course or courses offered by the teacher. In other words, we can regard the effect of such a teacher as predominately reinforcing, that is, he strengthens in the student a positive attitude towards himself as teacher and perhaps towards the course.

The influence of teaching behavior, however, is not predominately reinforcing in all classrooms and in all situations. Some teachers produce negative feelings. Technically, we say that such instructors produce more aversive stimulation than reinforcement. When a person's acts constitute a negative, or aversive stimulus, we say that person has delivered some degree of punishment. Both reinforcement and punishment can exist in various degrees of intensity. The combination of those two forms of influence can be conceived as having a significant impact upon the way a student behaves.

The effects of reinforcement and punishment become somewhat involved and complex; and studies of their influence upon behavior have shown that the results are not always in accordance with the obvious and simple predictions that can be developed from the theory. For example, if a child has been controlled by his parents largely by threat and punishment, he apparently must go through a phase of adjustment before he can get used to the reinforcing kind of treatment. Furthermore, it is not yet clear that strong reinforcement should be used exclusively in teaching, particularly when such reinforcement comes mainly from the teacher. It is conceivable that strong positive feelings by the student toward the teacher could introduce a kind of "emotional noise;" that is, a condition that could lead the student to accept uncritically the content provided by the teacher. And if we accept the proposition that critical listening and critical thinking are desirable habits to promote in school, it seems possible that the teacher can mis-manage the use of positive reinforcement.

On the other hand, the teacher who has a consistently aversive effect on the student may damage learning even more than the teacher who is consistently reinforcing, the child is likely to invent ways of escaping what the teacher says by turning his attention to irrelevant tasks, such as byplay with other students, or daydreaming. Also, too much aversive stimulation (punishment) can probably reduce the effectiveness of communication when the student wants to learn the content presented by the teacher. Negative emotions may thus produce a situation that results in a student's rejecting or failing to accept the valid information given by a teacher.

The responses to reinforcement and punishment are likely to be deeply rooted in the child's early training. As indicated before, the mother is associated, in the child's experience, with the presentation of a wide variety of reinforcers (those things or processes that result in reinforcement). And the infant comes to value her attention and affection in the course of his reinforcement history. Later in life, the infant learns that other people are less dependable than his mother in meeting his needs and demands. A basis is therefore provided for establishing a crude but effective way of sorting people into two classes: those who provide positive reinforcers and those who do not. When new acquaintances are made, the child is likely to judge them on the basis of similarities noted in other persons who have dealt with him according to the kind of treatment he has experienced from them. The extreme, clearcut preferences so formulated are likely to produce emotional tinges. Nevertheless, these preferences are products of overgeneralization; they are quite often derived from invalid assumptions. The tendency to generalize thus remains despite its being at cross purposes with the learning of necessary differences between stimulus objects. Side effects from negative feelings, for example, may remain for long periods, influencing the child to keep a safe distance between himself and those people, objects, and situations, which are seen as aversive. Unfortunately, in maintaining that safe distance, the child reduces his chances of learning more useful ways of deal-ing with his environment.

Carrying his complex tendencies with him into the school, the child faces new adult figures in strange settings. Each time he meets a new teacher the child is disposed to form some kind of feeling about the teacher. If the emotion is negative, he may demonstrate his lack of acceptance with responses that range from such mild ones as slowness in complying with requests, to such intense ones as crying, kicking, screaming and fighting. Unless the teacher can handle these problems skillfully, a mutual negative relationship may result and the attitudes remain unbroken. Fortunately, an awareness of disruptive factors in the school setting can be taught, and once learned, may serve to improve effectiveness in learning.

There is little agreement concerning the measurable specifics linked with good teaching. Uncertainty about the nature and measurement of teacher personality and about the relations between teacher personality and teaching effectiveness has been deplored by Getzels and Jackson (18), who cited over 150 references on the subject. In what light should a teacher trait be examined? Where should the making of a hypothesis concerning teacher characteristics and their predicted effects be found? The Committee on the Criteria of Teacher Effectiveness of the American Educational Research Association suggested that they ought to be found in learning theory or social-psychological theory, or even in any other body of theory in which a meaningful basis for developing the hypothesis exists. Both the relationships of teacher traits to student responses are often inferred directly from test scores, anecdotes, subjective ratings, or even from emotional reactions, without connecting these data with carefully developed hypotheses drawn from an adequate theory. The result is nothing short of chaotic, because of the questionable reliability and validity of the instruments used for gathering the data. While measuring instruments remain primitive, there must be even heavier reliance on theory for the meaningful examination of results and for the laying of a firm groundwork for later research.

TEACHER ATTITUDES

A study of teacher attitudes was conducted by Leeds (32), who used the Minnesota Teacher Attitude Inventory to measure attitudes associated with teacher-pupil relations. The instrument was used to obtain scores for predicting the ability of the teacher to get along with pupils in interpersonal activities. In general, he found that teacher-pupil relations were correlated with those teacher attitudes measured by the inventory.

Callis (12) explored the relation between teaching experience and attitudes expressed by teachers. His results showed that the first six months of professional training did produce significant changes in 20% of the attitudes measured. Some of the training effects were nullified, however, by only six months of on-the-job experience. Significant changes were produced in the undesirable direction in 11% of the trainees after they became regular teachers.

TEACHER INTERESTS

It would seem that teachers having strong interests in performing social service ought to have attitudes toward teaching that facilitate learning. While it has not been shown that the permissive atomsphere is the only cradle of all good things academic, or that the authoritarian is all bad, still there is a general acceptance of the idea that a combination of social interests and a moderately permissive attitude is desirable for a teacher. Beamer and Ledbetter (8) used the Minnesota Teacher Attitude Inventory and the Kuder Preference Record to test whether 164 experienced teachers held social service interests which would correlate positively with their attitudes. They found that many persons engaged in teaching did not exhibit two of the traits considered important for teachers, namely, interest in social service and a permissive attitude toward children. These results suggest that many teachers may not be predisposed to provide the kind of reinforcement and support needed by students to encourage their efforts.

Much has been written about transfer of learning, the logical defense for most if not all educational goals. But transfer has been applied almost entirely to the learning of factual information, skills, and general principles—all cognitive aspects of learning. But the idea of transfer can also be applied to the learning of attitudes and feelings. Teachers who influence students to fear mathematics, for example, may be the source of a lasting effect that could inhibit the development of an important potential.

TEACHER-PUPIL RELATIONS

If teacher-pupil relations are viewed as lying along a continuum, at one extreme may be found a positive pole, representing good rapport, warmth and acceptance, and at the other extreme, a negative pole, representing resistance, non-acceptance and dislike of pupils. In the positive direction, teachers are characterized by ability to arouse positive feelings in pupils. In the negative direction, a very different sort of teacher typically arouses negative feelings. Most teachers are found at neither extreme.

The positive teacher may not only influence pupils to accept many of his attitudes, values, and even his entire personality, but may also persuade the pupils who listen eagerly to his side remarks and trivial statements to assign an importance to them that is exaggerated beyond the teacher's intention. This results in an uncritical acceptance of many communications and in little attempt at verification by the pupil. For such pupils, the teacher can make no wrong pronouncements, and he is often quoted as a final authority by his pupils on the basis of some casual remark, even in the face of contradictory evidence.

A negative teacher, on the other hand, will stimulate, in some pupils, efforts to avoid the tasks at hand and perhaps even evoke rejection of some messages that are valid. Also, it is likely that some pupils will react only superficially to important information and will fail to learn the proper emphasis that the teacher intends. In general, the feeling tone that develops in the negative stiuation can act as a disruptive or inhibitive influence on learning. Under the positive teacher, however, a different, but no less real disturbance can result—a clouding of clear understanding through an overgeneralization of acceptance.

Fortunately, it seems that many pupils acquire sufficient objectivity to avoid too much spread of feeling tone from the teacher to subject matter. Yet, it is probable that many others, those most in need of help, suffer interference with learning due to arousal of extreme attitudes and feelings.

Strong emotional involvement may have a twofold influence. The first amounts to an interference with learning during its progress. Interference with the process of learning means that the learner fails to make important discriminations, fails to grasp the general meaning of the content presented to him, fails to make the proper interpretation of certain key words—any one or all of these learning errors could be influenced by lack of readiness, aptitude, and background experience, and by "emotional noise" in the situation. The second influence results in a transfer of a dislike of the teacher to a dislike of the subject matter, in the case of negative feelings, and in a transfer of acceptance of the teacher to warm and enthusiastic acceptance of the subject matter, in the case of positive feelings. The transfer of feeling tone from the teacher to the course content is likely to be cumulative throughout the school term, or for as long as the pupil persists in his feelings.

Whether learning is regarded as a particular process, or only as an inference from a change in behavior, it is important to uncover the extent of the influences of emotional arousal on cognitive learning. Although the behaviors of knowing and feeling may be as inseparable as nature and nurture, those two large response classes can be reasonably well identified and differentiated one from the other. As Mowrer (3) indicates explicitly, and as Skinner (4) reluctantly implies by his concept of reinforcement, learning always seems to involve some feeling tone. But when an unusual amount of emotion is introduced by extraneous stimuli in the learning situation, the effectiveness of cognitive learning may be significantly reduced. The experiments reported in this volume were not designed to separate feeling tone and knowing in the core of the learning process, but rather to estimate the effects of different kinds and amounts of emotional stimuli within the learning context but extraneous to the central messages to be learned.

The transfer of feelings from person to subject matter has been reported by Newcomb (37), who wrote about "strain toward symmetry." In moving toward the removal or reduction of conflict in one's psychological world, several decisions are possible. For example, a teacher, in trying to resolve conflict, may tend to develop the same orientation towards a situation as is held by a student. Or, he might attempt to change the student's perception, or even to distort it. Mutual negative feelings might grow as the teacher gradually perceives mounting dislike toward him expressed by a student. On the other hand, if the teacher is regarded as positive, the student is likely to change his orientation in favor of agreement with the teacher. Burdick and Burns (11) tested the power of Newcomb's speculation and they found that persons did change their opinions toward agreement with another who was liked (the positive valent other). Subject matter and its implications were preferred when they belonged to a liked or respected person. Hovland (22) also found that a positive source, such as a respected teacher, needed only to advocate an opinion in order for it to be accepted by students.

What the learner brings with him in the way of attitudes and past learning is sometimes at odds with what he is expected to accept. Uncritical thinking may sometimes be promoted by situations that appear incongruous or incompatible to the student. Bettinghaus (9) defined an incongruous attitude toward a topic as one that was at variance with a subject's attitude towards a speaker and/or his delivery. Results of the study by Bettinghaus showed that persons tended to achieve a balance between their attitudes towards a speaker's delivery and towards the speaker himself.

Perhaps the most substantial generality supported by the literature is that emotional stimulation, as a broad component in communication, has a significant influence upon learning. It is with this generality as a background plus the earlier speculations in the chapter that the rationale for hypothesis making is presented.

The disruptive effects of emotional stimuli have long been noted in a gross fashion, but efforts to measure the levels of arousal in learning experiences have been meager. In general, two different aspects of arousal must be considered. The first is simply a change in a physiological measure during a period of research observation. An example would be a sudden but significant increase in the heart rate or in the rate of breathing. A number of such measures exist and are recorded by the lie-detector mechanism, also called the polygraph. The second kind of arousal is a person's own verbal expressions of his feelings. Of interest here is not merely the person's verbal report of the name of an emotion that he presumably feels. The focal point is the reflection of the learner's attitude, or orientation, towards the situation in which he has become involved. Thus we have at our disposal an instrument measuring physiological changes, and which is therefore independent of a learner's formulated responses, and another instrument recording of the conscious feelings of the person, which does take account of formulated responses. In the experiments described in this report, the situation towards which a learner was oriented included both the instructor and the subject matter, the material that the person was expected to learn.)

The point of concern is the extent to which arousal will be followed by changes in learning effectiveness. The study was designed to introduce extraneous stimuli that could be judged as producing arousal, including changes in conscious feelings, so that relations between measures of these phenomena and learning could be determined. McGeoch's suggestion that cognition may be affected both by set (how the learner is predisposed to behave in a situation) and by the configuration of the whole context has been accepted by McKinney (35), who found that learning efficiency was reduced when emotion was induced in the learner. For McKinney, as for the present experimenters, all that mattered was the bare fact of arousal, and not any descriptive label attached to it. An experience with a snake, for example, may arouse what one person calls "disgust," while another may call it "fear," and both names may have the same behavioral manifestation. The disruptive effects upon learning may be similar, depending more upon the amount of arousal than upon the particular name assigned to it. Whether the learner views a negative instructor as "disgusting," "disliked," etc. is less important than the intensity of his reactions and the relation between the latter and learning effectiveness. The conscious ratings by students of teachers and subject matter ranged along several separate dimensions from strongly negative to strongly positive. We did not expect to ignore such ratings. They were used both to determine the direction and the intensity of conscious feeling and to provide a basis for an interpretation of the physiological arousals. Thus, if a student reported strong negative ratings of a teacher while showing many physiological arousals during the presentation of the negative teacher, we presumed that those physiological arousals at least accompanied negative feelings of a conscious sort. While the direction and degree of conscious ratings were considered important, any particular emotional word used by the student was considered too ambiguous to have a precise meaning.

Our research was not intended to unearth detailed causal elements. And no attempt was made to take advantage of presumed facilitative effects of emotion, such as determining the strong interests of the learner and building subject matter around those interests. Our attempt was to establish both positive and negative teaching conditions, determined largely by instructor personality. Secondly, we tried to get some measures of the degree of arousal produced during learning, including both the physiological and the conscious interpretations. Thirdly, we arranged to gather the student's evaluation of the teacher and of the subject matter after exposure to them. Fourth, we expected to measure the amounts of learning under the above conditions by means of tests of achievement. In effect, the question was: "To what degree is a learner aroused physiologically during an emotionally slanted presentation of subject matter, and how does that relate to both his evaluation of the learning experience and the amount that he learns?" Thus, the rationale for selection of hypotheses had to do with the interfering or disruptive effects of emotions on the learning of factual content, which was unembellished with emotional words. The emotion-arousing qualities studied were the teachers attitude, his bearing, his mien, and his mode of delivery. Teachers who were rated as strongly positive by students were interpreted as having a predominately reinforcing effect. The specific ratings were expected to suggest the focal points of reinforcement and aversive stimulation. In general, the resultant influences upon learning were expected to be in accord with Thorndike's Law of Effect.

DEFINITION OF TERMS

The words listed below require some explanation because of their special uses in the study.

Affect. The conscious feeling tone indicated by the student in an experiment, when he was asked to rate his teacher and the subject matter.

Arousal. The physiological and verbal responses of a student during the presentation of learning material and during testing periods. The physiological response was measured by an electronic mechanism described later in the report (page 18ff). Verbal ratings were collected by the use of semantic differential scales.

Arousal level. The term is defined operationally as the amount of change in electrical skin resistance recorded on a moving paper tape during the learning and testing sessions. It also refers to the number of galvanic skin responses recording during a session.

Artifact. An uncontrolled source of deflection recorded on a strip chart (the moving paper tape). An artifact may be produced by excessive hand movements of student in an experiment, by his squeezing of the electrode, or by inductive pickups of externally produced static discharges (seen on the chart as rapid pen deflections, or spikes). A retraced line on the chart is an artifact, caused by failure of the paper to sustain its movement. Spikes which are not quite so sharp may be produced by the operation of electrical machinery in the vicinity of the recording equipment.

Emotional noise. A hypothetical notion that refers to the disruptive action of external stimuli in the learning situation. The disruption is presumed to be caused by the influence of stimuli that produce emotional side effects which interfere with effective learning of material.

Gain. In this study, gain refers to the setting used on the input equipment of the amplifier (the circuit used to enlarge the tiny electrical skin response so that it can be recorded and made visible). Since gain represents a relative amplification, a corrective factor was applied whenever the gain setting was changed. Reasons for changing gain settings will be given in the procedure section.

Galvanic skin response wave. During the experimental period, a pen traces a continuous record on a moving page or strip chart. The pen's deflection from a straight line tracing is caused by changes in the electronic circuit by which it is driven. When a student is connected to the input of that electronic circuit by electrodes taped to his left hand (or right hand for left-handed students), the skin surface between the electrodes permits a very small electric current to flow. The amount of current flow is controlled by the resistance of the skin. When a student is aroused, his skin exudes a small of amount of sweat, which facilitates the flow of current, causing a noticeable deflection in the pen recording. The drier his skin, the more it resists the flow of current. Change in the amount of sweat secreted is presumed to have a connection with some emotional change, although the precise descriptive facts still remain undetermined. A galvanic skin response wave (hereafter abbreviated as GSR wave) is herein defined as a pen tracing the duration of which ranges from 1.0 to 99.9 seconds and whose amplitude may be from just above one millimeter to the entire width of the paper; that is, to a lateral sweep of about five inches. The wave is not merely the result of a reflex, such as a startle, but may be due to a slower physiological response, such as might result from normal fluctuations in stimuli from the outer environment.

Mode of presentation. The means by which the subject matter is transmitted to the student refers to mode of presentation. Three modes of presentation were used: audio tape, sound film and the teacher himself.

Negative transmission. When the subject matter is presented or transmitted in such a way as to provoke aversive feelings in the student, it is called negative transmission. Whether or not an instructor provokes negative feelings is determined by the way he affects the learner.

Positive transmission. A presentation of subject matter that results in an accepting and warm feeling in a student.

Semantic differential. A seven-point rating scale used to determine the conscious feelings of students about both the subject matter and the instructor. This scale is described more fully below (page 17).

Sensitivity. The relative ease with which the electronic recorder responds to changes in electrical inputs. In the present application, sensitivity refers to a position of the sensitivity control knob. One-volt and ten-volt settings were available. At one volt, the recorder was more sensitive than at ten volts, so that a given change in input produced a deflection ten times greater than was produced at the ten-volt position.

Servo-recorder. This instrument is sometimes called a pen recorder. It consists of an electronic amplifier which increases the magnitude of voltage variations fed into it, and then applies the power that corresponds to these variations to the movement of a pen. The pen is deflected laterally on a continuous page which moves at a constant speed. (See Plate No. 1)

Skin resistance. When a voltage source is connected across a pair of electrodes, and these electrodes are placed in contact with separate points on the skin, the skin introduces resistance to the flow of electrical current. This skin resistance is often measured in thousands of ohms, or K-ohms, or simply K.

Strip chart. The roll of graph paper on which the servo-recorder pen tracings are made is called a strip chart.

THE HYPOTHESES

The following expectancies, based on the rationale given earlier, form the immediate grounds on which hypotheses were made: (a) Differences between individual patterns of arousal and the conscious feelings induced by the teaching situation will cover a wide range because of differences between persons in their histories of learning and reinforcement. If such variations are found, certain predictions of group performance can be based on relations discovered in the study of the personal patterns. The following prediction serves to clarify this point. (b) Criterion test items of cognition based on those segments which just follow deviations of strongest affect will contain the highest error rate. Strong emotional stimulation experienced just at the conclusion of a meaningful message may serve to protect its retention. The information which just follows the emotive crest, however, may be masked, especially if there is a break in the continuity of meaning. The reason for that expectancy is a presumed trace of affect (overflow of emotion) that masks clear reception of new information. Error rate will be high when the masked content contains a shift in continuity, such as the introduction of a new paragraph just following the strong emotional response. (c) A gradient of affective reactions will be found in the majority of individual patterns, and it will be systematically related to the intentional control of emotional stimulation.

The following hypotheses are intended to be consistent with the above expectancies and with the rationale previously indicated.

a. Arousal is a function of the mode of presentation. There were three modes of presentation: audio tape, sound film, and the teacher himself. The hypothesis is based on the idea that the number of active stimuli is greater when film is used than when only audio tape is used because visual stimuli are added when a message is delivered by film. Also, the number of active influences on the learner is greater when a teacher is present than when a film is shown because of the dimension of depth and the sense of immediate contact. But the essential stimuli (those that carried the information to be learned) were carried by voice in all three modes of presentation. It is

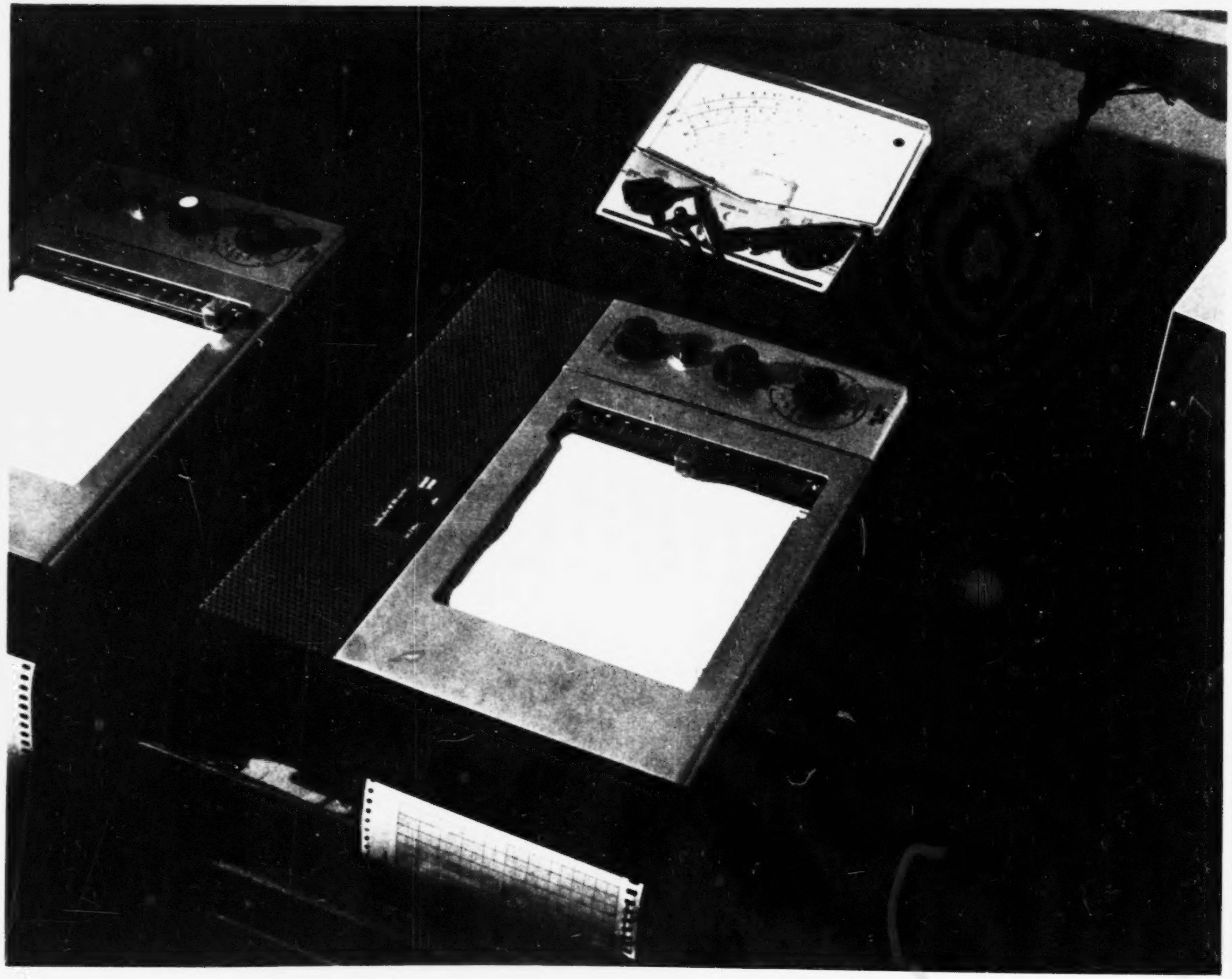


Plate I. Bausch and Lomb Servo Recorders for Making GSR Tracings.

therefore believed that, with respect to the amount of extraneous influence present, messages given by tape contain the least, film is intermediate, and presence of a teacher contains the most. Consequently, the hypothesis predicts that the amount of arousal (measured from both physiological responses and conscious feelings) will form a gradient, descending in this order: live, film, tape.

b. *Under the negative teacher, achievement scores decrease from tape to film and from film to the live condition*. Negative affect is presumed to produce greater distraction than positive affect, particularly as reflected by scores on factual test items. And because the number of emotion-producing stimuli increases from one mode of presentation to another (as indicated above), the result will be a diminution on effective learning.

c. *Under the positive teacher, achievement scores follow the same tendency as under the negative teacher, but the gradient will be more gentle than under negative presentations*.

d. *The strength of conscious ratings made by students of teachers increases as the mode of presentation changes from tape to film and from film to live delivery*. The negative teacher will be rated lower on film than on tape, and lowest in the live situation. The positive teacher will be rated highest in the live situation and lowest on tape, with film in between.

LIMITATIONS OF THE STUDY

The generality of the conclusions drawn from this study is restricted because of the nature of the samples, which are described in the next chapter. Both college and high school students were studied, the former from The University of Michigan, and the latter from The University High School, Ann Arbor, Michigan.

Inferences based on arousal levels do not necessarily apply to the broader areas of attitude, set, or specific emotions.

The brief learning sessions were not comparable to a regular classroom session in duration. Consequently, it is difficult to compare the experimental conditions with the typical teaching-learning atmosphere that obtains in ordinary classrooms.

When experimental controls are introduced to narrow the possible effects on a measured outcome, the number of variables that ordinarily operate in the field setting is necessarily restricted. Although such controls are invaluable for determining functional relations between presumed causes and effects, they limit the similarities that can be drawn between the experimental situations and the less controlled more natural settings provided by ordinary classrooms on which nevertheless the experimental situations are modeled.

Because the procurement of students for the experiments is likely to influence them to perceive the experiments as special duties which are not integrated with their regular pursuits, it is difficult to estimate the possible effect of the mental set so produced on the data. Although efforts were made to minimize that kind of probable bias, the researcher must proceed on the assumption that its effects have been minimized.

ASSUMPTIONS

All research efforts must rest on the acceptance of certain untested statements. It is usually not feasible to state explicitly all assumptions basic to a study. The following list includes those deemed most important:

(a) that readings obtained from the tracings on the strip chart accurately reflected physiological arousal in the subjects of the experiments. The instruments were tested by means of the usual startle stimuli and found to be sufficiently sensitive.

(b) that the students did not intentionally distort the semantic differential ratings; that is, that they provided reactions which validly represented their conscious feelings towards the teacher and the subject matter. Any apparent lack of coöperation in this respect was noted.

(c) that the persons who were used as teachers in the experiments could project perceptible positive and negative attitudes in amounts sufficient for producing arousal differences, especially within the same student as he experienced both positive and negative teachers. Although the data bear upon this assumption, which was later tested by the evidence, it had merely to be assumed to hold at the beginning of the study.

(d) that the influence of artifacts would not significantly bias the data. A pilot experiment was run to determine sources of artifacts and other errors, and modifications were made to minimize their effects to controllable limits.

CHAPTER TWO: CONSIDERATIONS ON INSTRUMENTATION

Modern psychology has not yet developed into a science that equals the logical rigor, precision, and objectivity of the natural sciences. The nature of the phenomena at the core of psychology resists the complete victory of radical behaviorism over mentalism and other subjective schools of thought.

While techniques of measurement are often based on operational definitions that provide a degree of scientific respectability, most of the psychological dimensions used to establish classes of data are still somewhat vague. The basic units of measurement have not been standardized and cannot be exhibited as concrete entities in a bureau of standards. Despite the difficulties of measuring psychological phenomena, the social importance of the concerns that psychologists pursue cannot be gainsaid, nor can it be denied that great strides have been taken towards the prediction and control of many forms of animal and human behavior.

The present study was not designed to introduce any new or even any improved methods of measurement. Consequently, the limitations of the data-gathering techniques used in similar studies in the past also apply to these studies. Nevertheless, an effort was made to manage the procedures for collecting data so that sources of error would be minimized.

The description that follows of the instruments and techniques used is deliberately extended, in order that the reader may be informed of the technical considerations that were necessary to gather the evidence for testing our hypotheses. This chapter may be omitted by the reader who is not interested in the technology of measurement. In spite of efforts to minimize technical language, it was ultimately impossible to make our descriptions adequate without the use of a significant number of technical terms.

THE SEMANTIC DIFFERENTIAL

The semantic differential, as developed by Osgood (46), has been used to analyze meanings in verbal contexts. Noble (39), however, insisted that affect, or the emotive influence of words, was what was really being measured. It was suggested that a semantic differential, or rating scale, patterned after Osgood's model, should not be used for"depth penetration." Nowadays, an experimenter does not use a ready-made, standardized semantic differential, but instead builds one suited to his specific applications. Concepts to be rated and factors to be used in rating them are chosen according to the conditions imposed by the experiment. Yet each semantic differential has common features, based upon extensive experimentation. A great many scales have been listed that are heavily loaded in their ability to discriminate levels of orientation towards concepts to be rated. Hence, once the concepts to be assessed are listed, the factors most appropriate for rating them can be selected from word lists containing the most powerful discriminators.

The validity of the semantic differential has not been fully established since there is no single, specific instrument involved. Yet the application of the method may be subjected to tests of validity. Griggs(20) employed a semantic differential with favorable results when examining differences in the semantic space between a control group and an experimental group, in making judgments between "self" and "neurotic" and between "ideal self" and "neurotic."

The discriminative power of the seven-point scale has been questioned, and the evidence thus far has been inconclusive. Gulliksen (21) felt that a scale of fifteen points or more ought to be used. In his examination of the instrument's discriminability, however, he focused primarily on its use in the measurement of meaning, not on the emotive influence of words. His study of attitude measures with time did not show much correlation with his fifteen-point scale.

Weinrich (45) was one of those who suggested that "what the semantic differential measures is not the meaning, but chiefly the emotive influence, of words" (p. 189). This is in agreement with Klapper who found that intuitive, impulsive, emotional expressions were encouraged by the use of the rating instrument. He also found that it could be used in substitution for a question, as in attitude testing, and would therefore serve as a more sensitive instrument, requiring about the same response time as would the answering of a question.

Other instruments have been used that produce similar results, but the semantic differential was selected because of its ease of construction and because of the great amount of research in which it has been employed. For example, it was chosen in preference to Thurstone's seemingly equivalent interval scale and to Guttman's scale. Since these scales are somewhat complex, research effort can be reduced by using the semantic differential. The scale developed and used for the present study appears in Appendix A.

THE GALVANIC SKIN RESPONSE

This phenomenon is variously referred to as the GSR, the Psychogalvanic Skin Response, the PGR, the Electrodermal Response, and the EDR, as well as by its most familiar name, the Galvanic Skin Response. It was once referred to as a reflex, though this is no longer common. Whatever the term employed, it may refer either to the change in apparent skin resistance or to a record made of that change. Some experimenters have thought that the change in skin resistance was brought about by the intensity of affect arousal, while others have believed that the change merely represented a state of tension which was not to be further defined. Lacy (30) said that it represented openness of the subject to his environment. McCleary (36) did not discuss inner causality, but mentioned three theories attempting to account for the resistance changes. The first was a muscular activity theory, which held that GSR was due to bioelectric changes in muscles. The second was a theory of vascular change, holding that GSR was electrical activity which accompanied vasodilation or vasoconstriction. This theory, however, has not been supported by convincing evidence of causation. The third theory, that of secretive changes, stated that GSR was an activity of the sweat glands. This is the theory which McCleary believes best supported by his evidence.

Gopalaswami (19) held the theory that the GSR is not merely an index of emotion, but is an indicator of increased learning effort. He tested this idea by imposing a mirror drawing task. But instead of noting less GSR deflection with practice, as would have been expected with the subsiding of emotional excitement, he reported increased GSR activity, such as would be anticipated if increased learning effort was involved. Balken (7), with a different kind of learning task, failed to find any relationship between deflections and efficiency of learning. Lists of pleasant, unpleasant, and indifferently toned paired associations were learned, and results did not support the learning effort theory.

Studies concerned with learning and GSR include simple rote learning tasks and problem-solving tasks. Esper (16) found that resistance rose at first, then fell until a subject reached the learning criterion. Slow learners' resistances fell lower than did those of rapid learners. Kuppers (29) also attempted to analyze records of subjects engaged in problem solving. He found a stepwise ascending curve for intellectual activity, a horizontal wave curve for heightened emotional involvement during problem solving, and a descending curve for complete relaxation or lack of concentration.

Unfortunately, it may not be assumed, in the aforementioned studies, that the degree of concentration or task complexity alone constitutes the independent variable. It has been found that verbal content must be controlled, so that the words themselves are not too arousing, or else they may also act as stimuli and produce GSRs. Haggard and Jones (23) found that the GSR could be used to discriminate the various levels of arousal induced by words having affective value. Affectively loaded words such as those usually considered by society as "taboo" elicit strong GSR activity.

Affective tinges associated with verbal content not only tend to produce GSRs, but other associations formed in the past have also been assumed to produce affective responses. Cofer (13) used GSR in seeking correlations between past emotional associations as revealed on the MMPI. He found no significant correlations. But Cooper (14) used GSR and found a positive correlation between attitudes strength and level of emotional support. He found it difficult to compare accurately the GSRs from subject to subject, so he made comparisons on an intra-subject basis. Silverman (42) also found that effectively charged words evoked arousal to the degree that they had been psychiatrically judged most personally meaningful.

Rote learning experiments have been plagued with confounding factors. It is not enough to determine, for example, whether arousal interferes with or facilitates learning. Just as has been the case with other classical experiments in learning, the investigator must ask whether that which is learned is also retained. Kleinsmith and Kaplan (28) showed that high arousal in a paired associate learning task correlated negatively with amounts of immediate recall, but that the high arousal was also followed by strong permanent recall.

Some attention must be paid to level of intelligence, whenever the GSR is used. Apparently intelligence level is related to reactivity. Ringness (41) divided subjects into low, average and high intelligence groups. When he imposed classroom-like tasks, he found the brighter children to be more reactive. Here, then, is a possible confounding variable, in the study reported by Gopalaswami. Ringness also found girls to be more reactive than boys.

Further evidence that GSR activity may be associated with attitudes or with orientation, which is generally affective, comes from attempts to examine the validity of "lie detector" tests. What seems to be influential in producing deflections is not so much the fact of lying, but whether the individual is acutely aware of it. For example, Lykken (34) used GSR to predict possession of "guilty knowledge," that is, knowledge that could only be held by the experimenter and his subject. Cooper and Pollack (14) correlated GSRs and attitude scale positions as indicators of affect in connection with prejudicial attitudes. The Spearman rank coefficient was 0.82.

Reliability

Several factors influence the reliability of GSR measures. Room temperature is one of these. Edelberg and Burch (15) found that the relationship between GSR levels and temperature variation was not constant, and therefore that drastic changes in room temperature ought to be avoided. They also found that the BRL, or Basal Resistance Level, did not change in a linear fashion. Some ranges of resistance fluctuate more readily than others. Reliability is also influenced by an adaptation effect, according to Rachman (40). With repeated exposure to stimuli, subjects tend to reduce their arousal levels, as indicated by GSR recordings. Random noise factors, introduced by the measuring equipment, may also make it difficult to read GSR records accurately. These and other differences, such as those between equipment chains, may make it difficult to make reliable inter-subject comparisons. Humidity differences, from one test period to the next, are not likely to produce significant changes, unless the changes are extreme.

Equipment types

The type of equipment used will depend upon the physical conditions of application and the kinds of data required. Some laboratories consist of very small rooms, with insufficient space for large, rack-mounted amplifiers and display panels. Space limitation suggests the use of light, portable equipment, of a relatively simple type. Where only a gross indication of skin resistance changes is needed, such as with pre- and post-experimental period notations, no separate amplifier, other than one contained in the pen recorder, is needed. Or it may be that the experimenter merely requires occasional resistance readings, so that no continuous recording is necessary, and then the pen recorder may be eliminated. If so, only a slow-reading highly sensitive ohmmeter is needed. However, many applications require not only a continuous, permanent record of both rapid and overall changes in resistance but also an amplification of the changes recorded. In this case, the amplifier ought to introduce a minimum of its own fluctuations, and should be reasonably linear throughout its range of amplification and recording. Larger deflections of the recording pen are not obtained only by increasing the amount of current that flows across the subject. Edelberg and Burch (15) found that current should be limited to ten milliamperes per square centimeter, in order to avoid skin effects which would result in a distorted record. The current density is a function of electrode size and the total current allowed to flow. The preceding consideration alone makes the findings of some of the early experimenters suspect, since the damaging effects of excess current flow had not yet been measured.

Both AC, alternating current, and DC, direct current, have been used in GSR studies. When current flows in one direction only, and follows a path consisting partially of fluids, such as sweat, then there may a tendency towards polarization. This electro-chemical property of the GSR circuit can result in a decrease in reliability of records taken over an extended period. Since an appreciable time lapse is necessary for polarization to produce a significant change, the effect can be nullified by rather frequent reversal of current flow. Thus, some experimenters prefer to use AC across the subject.

Modern technology has developed sensitive electronic amplifiers capable of accepting minute changes and enlarging them for more convenient inspection. Usually a bridge type of input unit is interposed between subject and amplifier, so that the amplifier will retain its stability over a broad band of input characteristics. Sometimes, however, some sort of voltage divider input is used. Whatever the type chosen, the output of the amplifier should be reasonably linear, since it must amplify across a range varying from about 10,000 ohms to 200,000 ohms. Some extreme cases will lie outside that range.

Too great an amplifier gain will produce the superimposition of extraneous changes upon the GSRs and upon the more gradual indications of resistance change. There are also random "noises" introduced by the amplifier itself, if the gain is too great. High gain amplification may also drive the pen off the recorder chart, unless some kind of automatic reset mechanism takes control when wide swings are developed. If a reset is not used, then the pen is re-centered on the page manually, by means of a centering voltage called "DC offset." If it is important to the experiment that the graph be accurately calibrated, then each time the pen is re-centered, the resistance reading should be noted and marked at the point of re-centering.

If the chart recorder is not calibrated for taking exact resistance readings, a sensitive ohmmeter may be used in combination with the amplifier-recorder chain. The ohmmeter may be used for visual readings, perhaps taken before, during, and after a period of stimulation. Such measurements provide an indication of levels of change of resistance, instead of the brief, sharp rises and falls which have been called GSRs. Sometimes the BRL, or Basal Resistance Level, is of interest to the experimenter. Burch, Childers and Edwards (10) speak of a GSR Wave, and their automatic GSR analyzer could be set to record onset-to-peak or "rise times" of waves with durations of up to 99.9 seconds.

Electrodes

It was once thoughtthat as long as an electrode was a good conductor, it could be used to study resistance changes in the human skin. Tarchanoff (43) used clay electrodes applied to cotton pads soaked in a saline solution. Fere' (17) seems to have used zinc plates. Most experimenters since that time have used zinc, copper, or silver. Livonian (33) used stainless steel. Treated silver electrodes do not corrode easily, and they may be used dry, though more satisfactory results are obtained when they are used with a good electrode paste.

The area of contact and the stability of contact between electrodes and the skin have been subjected to considerable investigation. Veraguth (44), one of the GSR pioneers, asked his subjects to grasp steel and nickel electrodes in their hands. Variations due to differential contact of the electrodes with skin surfaces surely would have been considerable. Direct current seems to have been used more widely in early research, so that one connection to the skin constituted the cathode, while the other became the anode. In 1925, Wechsler thought that the size of electrodes was a matter of convenience. Because of the importance of current density, however, electrode contact area may be critical. A change from large to smaller electrodes, for example, may necessitate a reduction of total current flow. Excessive current density may not result in any discomfort to the subject, but may still distort results. An advantage obtained from using large electrodes, with DC, is that there will be less polarization, due to the lower density of current, and hence a lower electromotive force will be acting as a confounding variable beside the measured changes attributable to autonomic arousal.

Choice of a site for electrode connection depends upon convenience to the subject and to the experimenter, and upon the desired sensitivity of the measurement. If experimental conditions are to approximate a classroom, it is obviously impractical to attach electrodes to the feet or to the chest. Sometimes, as was the case with Livonians' (33) study, attachments to many subjects must be made, and each subject must be taught to make his own attachments. Electrodes may then be designed which can be slipped on the fingers like rings. The use of paste would also be contraindicated, and the electrodes themselves would be dry.

Several palmar connections are common. Electrodes may be connected between two fingers on the same hand, between a finger and the forearm, or between the palm and the forearm. Some combinations are more sensitive than others, and some are less subject to artifacts of movement. If a palm connection is merely taped on, a fidgety subject may easily disturb the connection, producing intermittent changes in resistance which usually show up as very sharp "spikes" on the record. Or the tape may simply work loose, so that a high resistance contact results. If a ring electrode is placed on the third finger and a cup electrode taped to the index finger, a good stable connection results, but the tape used must not be so loose as to permit disturbance by accidental movement, nor so tight as to constrict the circulation.

Electrode Paste

Two considerations seem to contraindicate the use of paste. When it becomes necessary for a subject to fit on his own electrodes, the use of paste is not acceptable. The composition of the paste available may also determine whether an experimenter will use it. A paste that dries quickly and thereby quickly changes its capacity to conduct current will introduce error. Slow-drying pastes which permit excellent contact have recently been developed. No polarization is present provided that these pastes are used with alternating current.

Artifacts

Sources of error may be recognized and discounted, they may be compensated for, or they may be eliminated. One such source is caused by movement of the skin areas which are in contact with an electrode, resulting in changes in the recorded resistance. In human subjects, that source of error is easily eliminated, since humans will usually follow instructions to remain still, or accept various props which will help them to remain so. Careful electrode attachment will also minimize movement artifacts. Inductive electrical pick-ups in wire leads sometimes produce "spikes," or extremely rapid changes which show up on the record. Sixty-cycle hum is also easily recognized, by its periodic nature, and can be eliminated by electrical shielding of wiring and by filtering. Sharp pops may be introduced by inadequate insulation of the lead wire near its connection to the electrode. Reinforcing the insulation up to the electrode will eliminate this. It has also been found that scratched or scarred skin will effect the GSR record, as will soiled skin areas. Polarization effects, when DC is used, may be eliminated by applying AC across the subject, then converting this to DC for amplification and recording. The entire equipment chain, from electrodes to recording pen, should be tested for drift (slow, uncontrolled change) and also for more sudden shifts which might affect interpretations. If paste is used for one set of readings and not for another it may be impossible to compare the records with confidence.

Other Sources of Error

Two good reasons have been offered for not beginning the period of stimulation immediately after connecting the electrodes. The subject needs a rest period, following the physical activity of arriving at the testing room. Some subjects may also need time to become accustomed to the novelty of the situation and reassure themselves that no shock is involved. While there are usually practical limitations upon how long subjects can be fitted before being stimulated, it is mandatory that no attempt be made to correlate stimuli and responses until a settling down of the resistance gradient has occurred. There is likely to be a large initial change, occurring in the first three to five minutes, which appears regardless of whether there is intentional stimulation or not. Such a change must not be attributed to the effect of the independent variable.

Few experimenters have much to say about soundproof or acoustically treated testing rooms. One reason may be that subjects become accustomed to the usual sounds of living, and therefore learn to shut them out. The hum of motor traffic, and even the occasional bursts of motor scooter acceleration and the like, seem to produce little change in the records of subjects who have lived on a college campus. Subjects may show startle when a telephone rings in the same room with them, but it has been noticed that the ring heard several doors away does not produce a significant response. The same can be said for footsteps heard in a hallway outside the testing room. It is important, however, that the participating subjects not be allowed to talk after the stimulus period begins. If a subject coughs, sneezes, or laughs,this should be noted on his record next to the sudden pen deflection it produces. It may also result in significant deflections on the charts of other subjects in the room at the time.

Units of Measurement

Both intra-subject and inter-subject analyses are found in the literature on GSR experiments, and the writers have sometimes found it necessary to convert the basic data into units appropriate to their analyses. Conductance, the reciprocal of resistance, is sometimes preferred, and its units are mhos. Resistance is perhaps more commonly used for direct measurement, and its unit is the ohm. Insofar as movements of the pen are concerned, if a swing to the right corresponds to a decrease in resistance, then it corresponds to an increase in conductance. Lacey and Siegel (31) recommended change in conductance as the convenient and appropriate unit for GSR measurements. Others have advocated various score transformations, so that inter-subject comparisons may be made. While non-linearity has been found within the resistance range covered by GSR activity, Haggard (23) found that changes in the general level of resistance need not affect the size of individual GSRs. That is, the relatively rapid rise and decline of a GSR may be superimposed upon a baseline which is itself inclined and non-linear; but the amplitude of the GSR is comparable with another baseline which may be produced in another resistance range. In some analyses, it may be desirable to use a unit which is independent of general level of resistance. Niimi (38) accepted the conductance measure (mho) as useful for measurement of general, or basal, level. However, he suggested change in conductance, percent change in conductance, and percent change in resistance as measures appropriate to GSR measurement. Other measures which have been used are log of conductance change and log of resistance change. Lacey's change of conductance was accepted as the appropriate measure in the present study, with the exception that the reciprocal, or resistance, was used.

CHAPTER THREE: THE PILOT EXPERIMENT

DESIGN OF THE STUDY

A small experimental classroom was set up that allowed for the testing of four students at once. Each student sat at a special desk (see Plates 2 and 3), which was wired so that two small electrodes could be attached to the student's non-writing hand, leaving his writing hand free to respond to paper and pencil tests.

All electronic equipment used for calibration, amplification, and recording was placed in a room adjoining the experimental classroom, behind a closed door. (See Plate No. 4) Students could neither see nor hear the operation of the equipment.

In the experimental classroom there was a film projector, a screen on which the moving pictures were shown, and an eight-inch loud speaker located six feet above the floor in the left front corner of the room. (See Plate 5). A tape recorder was placed in the equipment room and the subject matter presented on tape was heard through the same loud speaker as the one connected with the film projector. Thus, provisions were made so that subject matter to be learned could be presented by three means: by tape, by sound-on-film, and by a teacher himself (live presentation).

Choosing teachers involved some preliminary explorations. We wanted the cooperation of three types of teacher: one that students thought of as predominately pleasing, another that had a neutral influence, and a third that evoked negative reactions in students. The results of our explorations revealed that the most adequate source of such teachers was a group of experienced actors enrolled in the graduate school and pursuing higher degrees in dramatics. The actors could effectively assume the intended roles, not only because of their special training, but partly because they had been exposed to the kinds of teacher that we wanted them to imitate. A selection of eight graduate student actors, recommended as the best by the head of the dramatics department, were given auditions before five judges. The actors presented the prepared content and the judges rated each actor on six semantic differential scales, having seven points per scale. After the auditions, ratings were studied to determine the selection of those best suited for our purpose. We found that differences in ratings between the positive, neutral, and negative speakers were highly consistent.

Tape recordings were made using the positive, neutral, and negative teachers. Ratings of the tapes agreed perfectly in rank order with ratings of the teachers in person, except the scores were less extreme than the live condition. We had anticipated that result in developing our rationale. Because of these encouraging outcomes, we believed that it was appropriate to

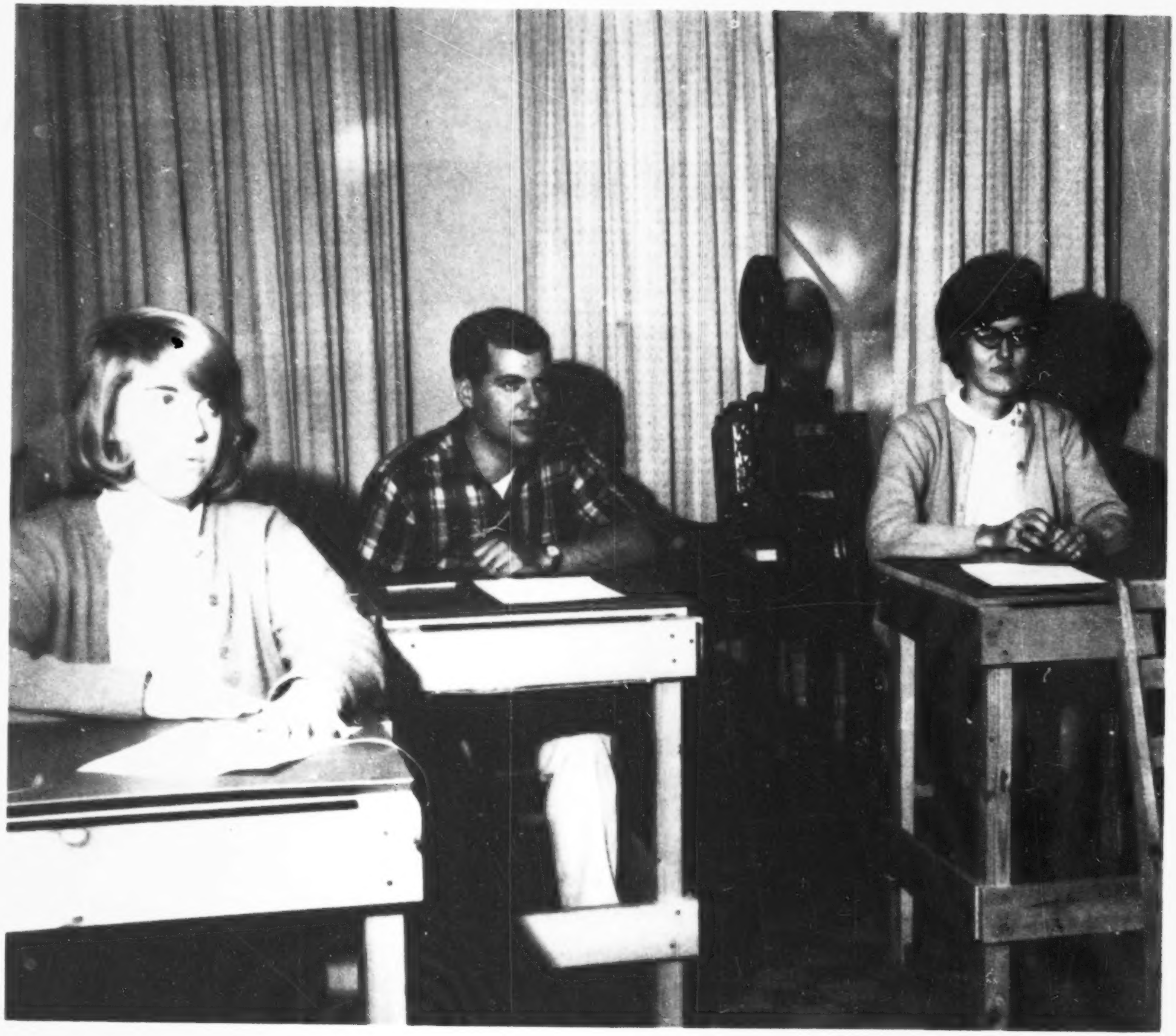


Plate II. The Experimental Classroom.

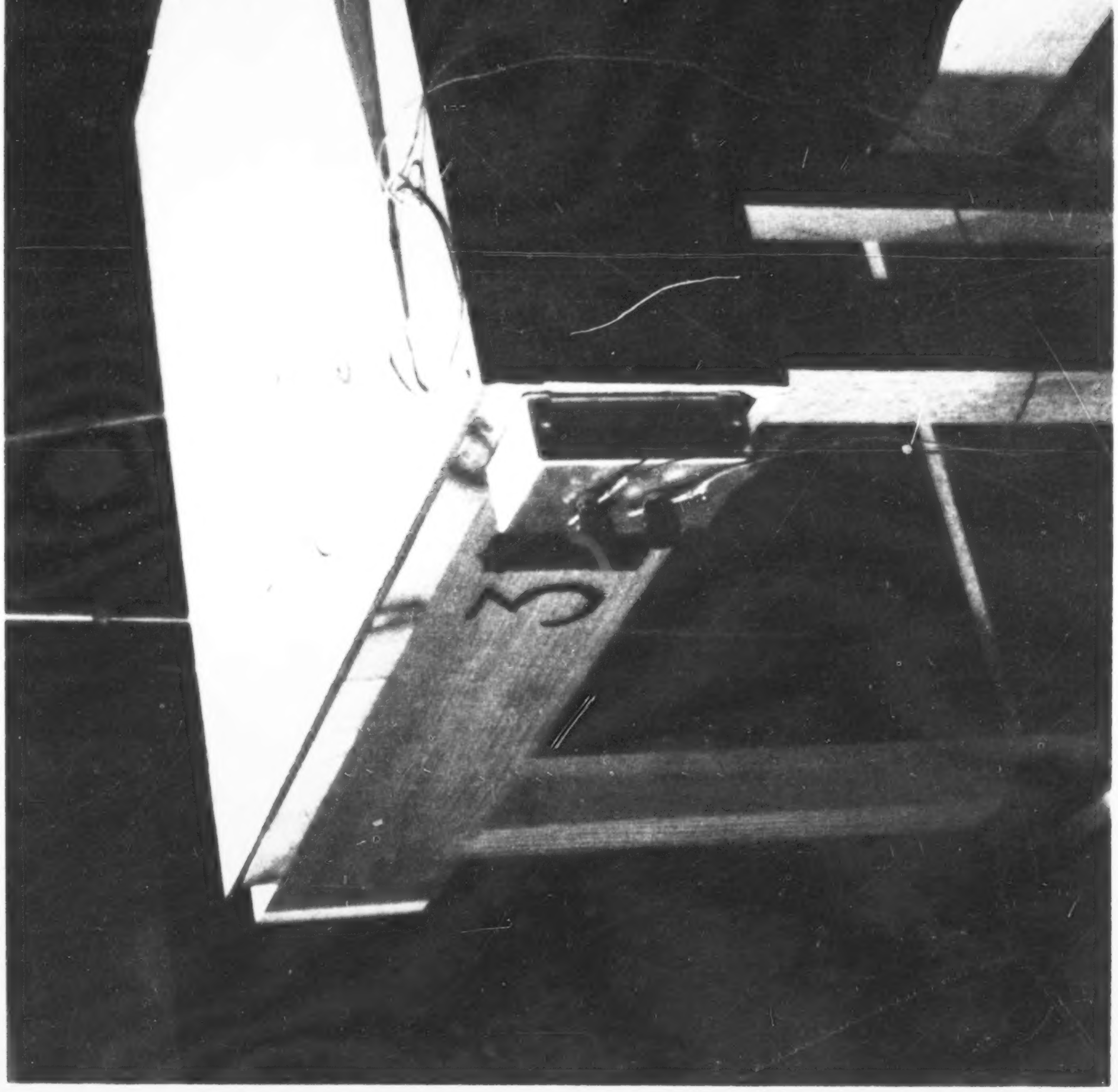


Plate III. Student Desk and Electrodes.

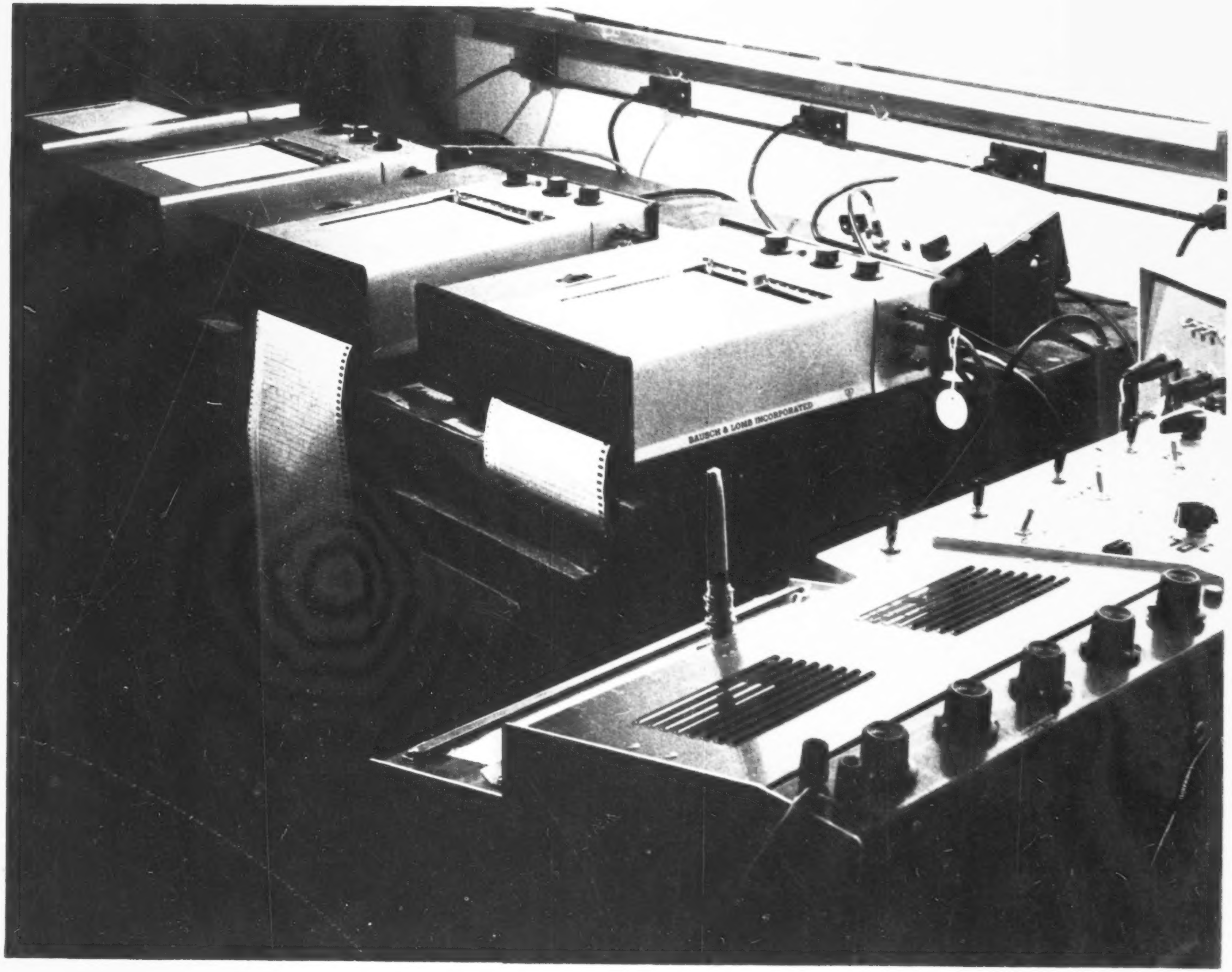

BAUSCH & LOMB INCORPORATED

Plate IV. Electronic Mixer, Amplifier and Recorders.

Plate V. Motion-picture Screen and Loudspeaker.

develop film presentations using the three "teachers."

Experimental subjects were drawn from a pool of college students enrolled in an elementary psychology course. All students in the course were required to participate in at least one research study as subjects. Twenty-six college students were used in the pilot study. Sixteen of the participants were freshmen, eight were sophomores, and four were juniors. There was the same number of men as women. In general, the students could be considered high achievers because of the screening practices used by The University of Michigan in selecting matriculants. The students used in the experiments could not be present without cooperating, because credit for participation could be granted or withheld by the experimenters. To make the requirement more palatable, subjects were paid a small stipend. All students expressed willingness to cooperate fully, and seemed interested in the experimental setup, which appeared to them as half classroom and half laboratory.

Films were produced by the Audio-Visual Education Center, The University of Michigan, under the technical direction of the Assistant Director of the Center. All films were made in black and white.

Although selection of students to be used in the experiment was not made by the use of a table of random numbers, or by any equally rigorous method of random selection, names were drawn from the list in an unsystematic way. The only restriction applied was that each sex should be equally represented. The names of the students were not familiar to the persons making the selection.

Nine messages were developed as the content to be learned by the student. Each message was approximately five minutes long as measured by the time required to present it on sound tape and on film. Two achievement tests were constructed for each message: one test covered the factual part of the message and consisted of 20 true-false items, and the other test was designed to measure how well the student could draw valid inferences from the subject matter for solving related problems. All inference tests were of the multiple-choice variety.

A schedule of experimental sessions was drawn up, and students selected for participation were notified in advance of the designated time and place. The schedule was developed with knowledge of each student's free time, so that conflicts were minimized.

At the time appointed for the experimental session, each student was conducted into the classroom and told what was expected of him. After electrodes had been attached and tested, students were told to listen carefully to the presented material and to learn the content as best they could, in order to be prepared for the tests that followed.

One of the four desks occupied by the students during data collection was equipped with an "emotional response mechanism," a device used for the recording of the positive and negative feelings of a student during presentation of the learning material. The student at the desk having the emotional response mechanism was exposed to the same messages, took the same tests, and was treated in the same way as the other three students during the session. The only exception was that this student was shown how to operate the mechanism and was told that, any time he felt positively stimulated by anything during the learning session, he was to pull the knob toward him (see Plate 6), and any time he felt negative about anything in the situation, he was to push the knob away from him. A slight pull or push meant a noticeable but relatively weak feeling, while a strong push or pull meant a strong feeling. The emotional response mechanism was wired to a Heathkit Recorder, which drew tracings of the student's responses during the session. A negative response was recorded on the strip chart as a movement to the left, and a positive feeling was recorded as a movement to the right. Thus, the record of feelings during the learning session could be gathered and permanently recorded for later inspection. The Heathkit Recorder was less sensitive than the Bausch and Lomb Recorders, which were used for the GSR tracings, yet the Heathkit Recorder was found to be quite adequate for its intended use.

One hour before each experimental period, all electronic equipment was turned on, calibrated, and tested for proper functioning. Just before the presentation of the learning material, and after each student had been fitted with electrodes, the initial skin resistance reading was determined for each of the four students, and the proper settings were made, so that the pen used to record the tracings would be centered on the strip chart.

After all preliminary tasks had been performed, the initial message was presented, followed by testing and a brief rest period. Then the second message was presented, followed by the appropriate tests. And finally the third message was given, plus tests. The entire session took about 45 minutes.

Messages were prepared in nine separate units: three were delivered by tape recording, three by film, and three live (the speaker in person). Each speaker, or teacher, assumed the same role during all his presentations. The negative teacher recorded a message on tape, a second message on film, and he presented a third one live. The same procedure was repeated for the positive and neutral teachers.

Tape recordings were produced by reading the script containing the content to be learned. Deliveries maintained a normal tempo that approximated both the filmed and the live presentations. The filmed deliveries showed the speaker and a blackboard in the background. About half of the filmed shots were close-ups, showing only the teacher's shoulders and head. Frequent use of close-ups provided for clear vision of the teacher, particularly of his facial expressions. About fifty percent of the shots revealed more of the surrounding conditions, indicating a typical classroom, as provided by appropriate props in the filming studio. Shots of the larger context provided an occasion for noting the bodily movements of the speaker, his posture, and what can be called his orientation towards his immediate surroundings. Each speaker had rehearsed his topic sufficiently so that he appeared to refer to the paper before him as notes that

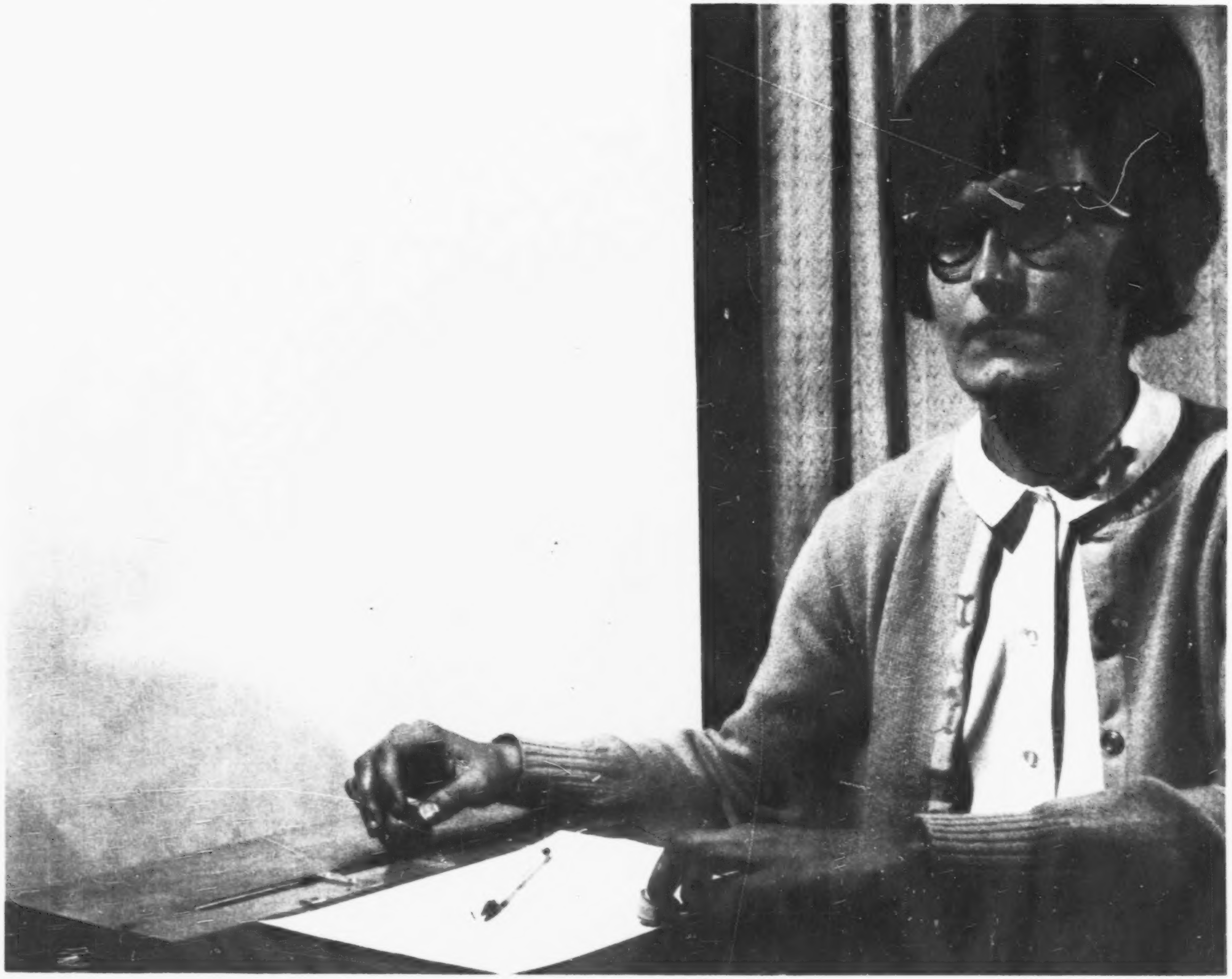


Plate VI. Device for Recording Affective Responses of Students During Learning Sessions.

covered a general outline of his presentation. Each speaker repeated the same role in the live deliveries. The main purpose of the live presentation, as compared with the filmed, was to increase the amount of extraneous stimuli and thus to intensify the emotional side-effects of the presentation.

The room in which the experiment was conducted resembled a small classroom more than it did a laboratory. A blackboard gave evidence of the room's earlier function, as did chairs of the usual classroom type. There were two partly filled bookcases. The desks, however, were not like the ordinary student desks. They were simply flat-topped tables constructed for the experiment, 20 inches square and 30 inches high. The left front leg of each table bore two small binding posts, to which GSR electrode leads could be connected. When the presentation was to be made live, the speaker stood near the loudspeaker, so that for all three modes tape, film, and live, sound originated from the same area, and was of about the same level of intensity. Walls and ceilings were not acoustically treated, hence GSR records were carefully watched during the sessions. But sounds from the hallway or street did not produce detectable GSR waves. The occasional passage of a particularly loud motorcycle produced no noticeable GSR effects. It was therefore reasonable to think that commonly occurring stimuli heard from outside the classroom had no biasing influence upon the data.

Subjects were requested not to talk, nor to move the electrode-equipped hand unnecessarily. In general, students complied quite satisfactorily with the request. Occasional movements of the wired hand were noted on the record to differentiate the GSR deflection from the others during a given session.

All testing of the group occurred in mild spring weather, with no radical temperature changes. Subjects were given some idea of the general nature of the experiment, and were put at ease about being attached to electrodes. Some of the students had already participated in experiments involving GSR measures.

Each session was scheduled for a group of four students. In cases in which not all students appeared for a session, those absentees were re-scheduled. Instruction sheets were provided, informing students in detail of what was to happen, and that they would be tested over the material to be presented. During the first session students were also told at the end of each segment or message the mechanics of responding to the true-false test, the inferential or multiple-choice test, and the two sections of ratings scales (semantic differential). Instructions were not repeated in their entirety after the first session.

No introductory remarks were necessary from the teacher for any presentation. In the case of live deliveries, the teacher entered the room, stationed himself before the class near the blackboard and began his presentation.

The positive teacher smiled often, appeared to enjoy his task, and maintained a pleasing manner throughout his delivery. He was a tall, clean-cut man, in his middle twenties, and produced the clear impression of having considerable experience in the teaching profession, as evidenced by his confident and easy manner.

The neutral speaker spoke clearly, but in a matter-of-fact manner, with no attempt at inducing any feeling tone.

The negative teacher frowned, used "uhs" and "ers" frequently, cleared his throat noisily, shuffled papers occasionally, and otherwise produced the kind of distractions that were less than pleasing. He did not ad lib or depart from the text. He was about the same age as the positive teacher but considerably shorter with less regular features. His dress was not conspicuous, but quite ordinary in appearance. He maintained a natural-looking smirk that conveyed the suggestion that he was not particularly interested in either the students or the subject matter, but that he happened to be there because it was expected of him. His articulation was clear, and there was no doubt that students could hear him without difficulty.

Inspection of the data after a few sessions indicated clearly that the neutral deliveries were not actually rated as neutral but somewhat as positive, yet considerably below the positive teacher's ratings. Consequently, data from the neutral presentation were not used.

The electrodes were made especially for this experiment by Coleman Enterprises. Specifications were similar to those for the electrodes used by Livonian (33), with the exception that the contact metal was chlorided silver. The contact areas (between skin and metal) were saucer-shaped so that the special paste used to maintain good contact could be contained in the saucer depression. The total contact area was so designed that not more than 10 milliamperes per square centimeter would flow across the area contacted. A plastic ring was fitted with pads, inside the ring, with one pad containing an electrode. The other pad was glued to the ring opposite the electrode so that a firm, gentle, non-constricting contact would be sustained. Rings were placed on the first and third finger of the non-writing hand. This left the writing hand free for movement without disturbance of the electrical connections. It was expected that every hand would contain some extraneous coating, so all subjects were supervised in carefully washing and drying their hands before being fitted with electrodes. Then the contact areas were washed with alcohol and the paste applied, both to the electrode cup and to the skin area. After the electrodes had been fitted, any paste remaining outside the contact area was removed. Also, the leads to the electrodes were inspected and cleaned.

The electrode paste was furnished by the Pharmacy Department of The University of Michigan and was of a type previously found to be satisfactory. It consisted of a suspension of sodium chloride, potassium bitartrate, pumice, tragacanth powder, propylene glycol, and water.

From the binding posts on each student's desk, a small cable led to an input unit, basically a Wheatstone bridge. Inputs were fed to a four-channel operational amplifier. Amplifier outputs fed four Bausch and Lomb servo-recorders. Chart movement was set at a speed of five inches per minute. Equipment was given a one-hour warm-up preceding experimental sessions.

As soon as the four students entered the room they were fitted with electrodes and given orientation. Five minutes or longer elapsed, during which the subjects relaxed and the equipment operators made a final check of the apparatus. There was typically an initial period of rapid resistance change, followed by a relatively steady state.

Recorders were allowed to run during achievement testing to estimate whether or not the testing periods produced more physiological arousals than learning sessions. Some students voluntarily wrote remarks on rating sheets, and were henceforth encouraged to do so.

Most students who were assigned to desk number 4, the one containing the emotional response mechanism, expressed a willingness to operate the knob which made possible a recording of their conscious feelings during presentations. As noted before, only one learning station was so equipped. The knob on the mechanism was attached to the arm of a potentiometer, which controlled the output of a dry battery. The output was connected to the pen of a servo-recorder, which kept a record of the number of times the student registered positive and negative feelings. Students were shown how to operate the mechanism and were told, "If you feel good about anything pull the knob toward you, and if you feel negative toward anything push if from you. The farther you push or pull the knob, the stronger the feeling you will register." They were told that this was their opportunity to express their feelings, but not move the knob unless they felt either positive or negative feelings during the learning session.

Table 1 below shows the assignment of messages to means of presentation and to polarities (positive, neutral, or negative teachers).

TABLE 1

ASSIGNMENT OF MESSAGES (UNITS) TO MODES AND POLARITIES

Polarity	Tape	Film	Live
Positive Delivery	Unit No. 1	Unit No. 3	Unit No. 5
Neutral Delivery	Unit No. 7	Unit No. 2	Unit No. 4
Negative Delivery	Unit No. 9	Unit No. 6	Unit No. 8

Note: Messages or units numbered 6, 8 and 9 were dropped because of elimination of neutral presentations.

Table 2 shows the kinds of measurements taken during the learning session and after it.

TABLE 2

MEASUREMENTS

	During Learning Session		Following Learning Session		
	GSR	Conscious Feeling	Factual Test	Inference Test	Ratings of Teachers and Content
Number of Students Measured	26	6*	26	26	26

*Only one student during each session could use the emotional response mechanism.

MEASUREMENTS

The specific measures taken from GSR records, graphs of conscious feelings, factual tests, inference tests, and ratings of teachers and instructors are indicated below.

(a) From the GSR wave patterns measures were taken of (1) the frequency of GSRs during each presentation, (2) rise time or the time required for a pen tracing to reach its peak from its beginning, and (3) the slope of the GSR curve during its rise.

(b) From the continuous graphs that recorded conscious feeling tone during each presentation only the frequency was counted.

(c) From the true-false tests two scores were determined, both the number of items correct and the number right minus the number wrong. The two scores were obtained to find out whether or not the use of a penalty formula gave data that were more usable than simply the score obtained by adding the number of correct responses.

(d) From the inference tests (multiple-choice), the score amounted to the number of alternatives that the student indicated were incorrect. One point each was given for every incorrect alternative that was so marked by the student. If, however, he happened to include the correct answer, a total of three points was substracted from the number of positive points. The kind of penalty formula used provided a means for obtaining rather wide degrees of individual difference with only a few test items. For example, the possible range of scores for a five-item test, having four options per item, is from a negative 15 to a positive 15.

(e) For ratings of teachers on the six semantic differential scales, a single composite score was derived. The maximum score for positive feelings that could be expressed by a single student was 42, and the lowest score was 6. Scores above 24 were considered on the positive side and those below 24 were interpreted as negative.

(f) For ratings of subject matter by students, the same procedure was followed as indicated in "e," above.

The various scores were tabulated for individual students and for groups by mode of presentation and by polarity (positive and negative delivery). Means and standard deviations were computed and tests of significance were performed to estimate the significance of the mean differences.

RESULTS

We predicted that students would be aroused more while learning under the live presentation than under the other two. We expected that arousal would be lowest when students learned via sound tape. Filmed presentations were predicted to produce a level of arousal between live and tape, but nearer to the live delivery because both contain audio and visual stimuli. Table 3 summarizes the data for physiological arousals, measured by frequencies of GSRs.

TABLE 3

OBTAINED MEAN SCORES OF GSR FREQUENCIES UNDER THE TREATMENTS

Mode	Positive Teacher	Negative Teacher
Live	3	5
Film	10	8
Tape	9	16

Results were almost exactly opposite to the expected trend. Our original rationale cannot account for the outcome. One possible explanation of the unexpected results is that under taped presentations, students felt disposed to concentrate more carefully than when visual stimuli accompanied the messages. And the greater concentration produced GSRs more abundantly than under the live form of delivery. Furthermore, it is reasonable to think that students expected to be aided by the visual cues. And as a result, the effort required to learn under filmed and live modes was perceived by students as easier than when sound alone was the only source of information. The notion that concentration itself produces physiological arousal merits further investigation.

Our results suggest that extraneous stimuli alone do not necessarily produce measurable GSRs, particularly when such stimuli emanate from the source of communication.

Table 4 shows the significance of the mean differences in the number of GSRs, based on data in Table 3.

TABLE 4

ANALYSIS OF MEAN DIFFERENCES BETWEEN GSR SCORES ACROSS MODES OF PRESENTATION (t-RATIOS)

Mode of Presentation	Positive Delivery	d.f.	P	Negative Delivery	d.f.	P
Live vs. film	2.12	24	.05	N.S.		
Film vs. tape	N.S.			2.92	24	.01
Live vs. tape	3.72	24	.01	5.69	24	.01

Table 5 is a summary of conscious arousal produced under the positive and negative teachers. The numbers represent mean ratings of the teachers on six semantic differential scales. To remind the reader of the nature of the scales, a sample is reproduced below.

/ 1 / 2 / 3 / 4 / 5 / 6 / 7 /
Unpleasant Pleasant

The student circled the number on the scale that most nearly represented his rating of the teacher between the two polarities. Ratings of the six scales were combined to produce a single total score from the responses of each student. Positive and negative teachers were rated just after each presentation: live, film, and tape.

TABLE 5

MEAN RATINGS OF TEACHERS ACROSS MODES ON SEMANTIC DIFFERENTIAL SCALES

Mode	Positive Teacher	Negative Teacher
Live	34	13
Film	31	15
Tape	30	17

Differences between ratings of positive and negative teachers within each mode of presentation was significant beyond the .01 level of probability.

The maximum possible score was 42 for each teacher and the minimum possible score was 6. A total of 24 represented a neutral rating, meaning that the judgment was midway between the polarities when all six ratings were totaled. Differences between ratings of positive and negative teachers were highly significant, beyond the .01 level of probability. Our interpretation of these results is that the teachers were rated by students in agreement with the intended roles.

It is worth noting that our first hypothesis, which predicted a decrease of arousal from the live through the taped modes of delivery, was supported by the conscious ratings but was opposed by the data on physiological arousal. Under the negative teacher, for example (Table 5), there is a clear trend toward the neutral rating, beginning with the live presentation and ending with the taped delivery. Differences between these ratings under the negative teacher, across modes of delivery, were all significant, better than the .02 level of probability.

Table 5 also shows ratings of the positive teacher across modes of delivery. These ratings also establish a trend toward the neutral position from the live through the taped presentations. Differences between means under the positive teacher were also significant at the .05 level of probability, or better.

There seems to be little doubt that the live presentations, by both positive and negative teachers, produced greater conscious arousal than the other modes of delivery, and that the predicted gradient for conscious arousal was supported.

Table 6 shows how students rated the subject matter under the positive and negative presentations by live, film, and tape deliveries.

TABLE 6

MEAN RATINGS OF SUBJECT MATTER UNDER POSITIVE AND NEGATIVE PRESENTATIONS

Mode	Positive Teacher	Negative Teacher
Live	29	19
Film	28	21
Tape	28	19

Differences between ratings under the positive and negative teacher within each mode of presentation with significant beyond the .01 level of probability.

The significant information in Table 6 is that the ratings of content approximated the ratings of the teachers. This information is interesting because all messages were quite homogeneous in content; they were all taken from the New Atlantis, which relates a voyage to the land of Orlu. The content from the New Atlantis was re-written to produce an easy and consistent reading level throughout. Each message was carefully tested for its autonomy; that is, each message could be understood by itself and did not have to be delivered in an order relating all messages. All messages contained a low and uniform loading of emotional words. Yet when the messages were presented by the negative teacher, they were perceived by students in a negative way; and when they were given by the positive teacher, they were rated well above neutrality. This finding supports the prediction that students tend to generalize their feelings from the teacher to the subject matter presented by him.

Table 6 does not establish as neat a trend in each column of numbers as does Table 5. Ratings of subject matter were all nearer the neutral position under each mode than ratings of the teacher, as predicted by the fourth hypothesis.

ACHIEVEMENT UNDER THE VARIOUS TREATMENTS

The second hypothesis predicted that, under the negative teacher, achievement scores would decrease from tape to film, and from film to live presentation. In other words, the predicted gradient of achievement was in the following order: tape (highest), film (middle), live (lowest). Our prediction was based on the idea that, under the stimulation of a negative teacher, students would learn best if the teacher were not visible; that is, when only such a teacher's voice was heard delivering the content on tape. We thought that the negative teacher in the live situation would produce the most emotional noise, and that this noise would interfere with effective learning. Table 7 summarizes the results that bear upon our second hypothesis.

TABLE 7

MEANS OF ACHIEVEMENT SCORES UNDER NEGATIVE TEACHER BY MODE OF PRESENTATION

Test	Live	Film	Tape
Factual	10.8 (5.1)	7.0 (3.8)	10.4 (4.0)
Inference	7.4 (4.8)	3.8 (4.8)	5.7 (4.6)

Note: Numbers in parentheses are standard deviations. They appear to be unusually large, because penalty formulae were used that increased the range of the scores. Some students actually received negative scores.

The second hypothesis is not supported at all by the evidence in Table 7. The data also have no apparent connection with the amount of physiological arousal, as measured by GSR frequencies. The filmed presentations were the least effective of the three, while the live and tape modes produced about equal achievement in factual learning. Scores from the inference tests produced exactly the same ranking order as the factual test scores, with respect to mode of delivery. From highest to lowest achievement, the order was: live, tape, and film.

Table 8 shows the results of analysis applied to the mean differences in Table 7. The t-ratios for correlated means were computed.

TABLE 8

ANALYSIS OF SIGNIFICANCE BETWEEN MEANS OF ACHIEVEMENT SCORES BY MODE OF PRESENTATION UNDER THE NEGATIVE TEACHER (t-RATIOS)

Test	Live vs. Film	d.f.	Film vs. Tape	d.f.	Live vs. Tape	d.f.
Factual	3.27*	24	2.90*	24	N.S.	
Inference	4.42*	24	2.39**	24	2.65**	17

* significant at the .01 level of probability.
**significant at the .05 level of probability.

The above table shows that differences in average achievement, as measured by factual tests, were significant in two out of three comparisons. Only the live-versus-tape comparison yielded no significant difference. All differences between achievement, as measured by inference tests, were significant at the .02 level of probability, or better.

Tables 9 and 10 repeat the above analysis for the positive teacher.

TABLE 9

MEANS OF ACHIEVEMENT UNDER THE POSITIVE TEACHER BY MODE OF PRESENTATION

Test	Live	Film	Tape
Factual	10.6 (5.1)	10.7 (4.0)	11.3 (4.0)
Inference	5.1 (4.8)	1.7 (5.9)	4.5 (4.6)

Note: Numbers in parentheses are standard deviations; they appear unusually large because penalty formulae were used that increased the range of the scores. Some students received negative scores.

TABLE 10

ANALYSIS OF SIGNIFICANCE BETWEEN MEANS OF ACHIEVEMENT SCORES BY MODE OF PRESENTATION UNDER THE POSITIVE TEACHER (t-RATIOS)

Test	Live vs. Film	d.f.	Film vs. Tape	d.f.	Live vs. Tape	d.f.
Factual	N.S.		N.S.		N.S.	
Inference	4.00*	24	2.69**	24	N.S.	

* significant at .001 level of probability
**significant at .05 level of probability

For means scores on factual tests under the positive teacher, the order of achievement from highest to lowest was as predicted; yet Table 10 indicates that there were no significant differences between achievement levels by modes on the factual test results. Analysis of the differences between means on the inference tests, as shown in Table 10, shows significant t-ratios in two cases out of three.

The principal motive for these experiments was comparison of the achievement levels of students experiencing both liked and disliked teachers. Our rationale suggests that the extreme feelings of like and dislike produced by teachers can both curtail effective learning. We thought that the positive teacher would produce interfering stimuli that would particularly affect learning above the level of simple acquisition of facts. Consequently, we expected reasonably good achievement under the positive teacher in the learning of facts, but we also expected a reduction in achievement level on inference tests. Unfortunately, it was not feasible to compare factual versus inference scores because basis of comparison could not be found. Another equally important comparison, however, was possible: between positive and negative teachers on each measure of achievement. Table 11 presents the results of this analysis.

TABLE 11

ANALYSIS OF SIGNIFICANCE OF DIFFERENCES BETWEEN ACHIEVEMENT SCORES UNDER THE POSITIVE TEACHER VS. THE NEGATIVE TEACHER

Difference Favoring	Factual Tests	d.f.	Inference Tests	d.f.
Positive Teacher	3.12	24		
Negative Teacher			3.60	24

Both t-ratios are significant at the .01 level of probability.
Note: computation is based on aggregate averages for factual tests, and on the aggregate average of all inference tests.
The actual mean difference for factual test was 4.40.
The actual mean difference for inference test was 5.70.

This analysis agrees with our original rationale. We said that a strong liking for a teacher would probably create a mental set incompatible with critical evaluation of the material. It therefore seemed probable that achievement measured by inference tests, which requires logical evaluation, would be a lower level under the liked than under the disliked teacher. The highly significant differences in Table 11 between the means of the cumulative averages of the three inference tests support our expectations. We predicted that a negative teacher would create the kind of emotional noise that would tend to interfere with acceptance of valid information. Consequently, it seemed reasonable to think that learning of factual information might be enhanced under a positive teacher. Table 11 also supports this prediction by showing a highly significant difference in factual learning in favor of the positive teacher over the negative one.

In a sense, the results of analysis in the above table are incompatible. It would seem that curtailment of factual learning would also curtail logical inference, when the latter depends upon the former. Yet, the important question is: How much factual information can be sacrificed without damaging logical problem-solving based on factual content? The answer is moot, because it would seem to depend upon the facts needed for answering any particular inference test item. In the present experiment, it seems that the negative teacher stimulated learning of enough facts to equip students to deal satisfactorily with problems requiring the use of logical inference. Further study of the problem is necessary.

One of the main weaknesses in the pilot study is that no tests of reliability were conducted on the achievement tests. These tests, however, were carefully made, re-written and edited repeatedly before they were used. The second experiment includes greater attention to achievement tests. Difficulty levels for each item were measured, discrimination indices were computed, and reliabilities were determined by the Kuder-Richardson technique.

Analyses in the pilot experiment that involve achievement tests must be interpreted with caution because of the weaknesses indicated above. We then proceeded with a second experiment with the belief that its results would be similar to those in the pilot experiment. Whenever possible, we attempted to rectify weaknesses found in the first study while designing and carrying out the procedures in the second.

SUMMARY

A pilot experiment was conducted to determine the effects of arousal produced by a teacher who was liked by students and by one who was disliked. We predicted that strong feelings, both for and against the teachers, would interfere with learning. In order to test this prediction, we made arrangements for the liked teacher to give his message in these ways: by tape, on film, and in person (live). The person selected to assume the role of the liked teacher was rated very high by students on a set of rating scales. We predicted that, because of the high ratings, the teacher would create a kind of emotional noise that would interfere with achievement, particularly of the kind requiring critical reflection. The disliked teacher used the same modes of presentation.

We predicted that when teachers delivered the content in person, the maximum number of stimuli would bear upon the student—those carrying the basic information, plus those carrying the extraneous stimuli due to the teacher's presence and behavior. We expected that taped presentations would create the least amount of arousal because the stimuli were reduced only to those conveying the subject matter. Consequently, we predicted a gradient of arousal according to the mode of delivery. The expected order from highest to lowest was: in person, film, tape.

Arousal was measured in two independent ways. The first used electronic equipment to record physiological arousals in the form of galvanic skin responses (the lie detector or polygraph method). The second kind of arousal was different: it involved the conscious feelings that students expressed towards the teachers and the subject matter.

Results gathered from the two kinds of arousal revealed that physiological arousal, measured by the frequencies of GSRs during each learning session, did not agree with our prediction. In fact, they were almost the opposite of those expected. The greatest amount of physiological arousal occurred when students listened to tape, and the least occurred when they were given information by the teacher in person.

Ratings made by students of teachers and subject matter agreed in general, however, with our prediction. Most of the neutral ratings were made for taped deliveries and most of the deviant ratings (both positive and negative) were made for live deliveries.

We found that students learned facts better from a liked teacher, but achieved more on inference tests under a disliked teacher. Differences in both cases were highly significant when total scores involving all modes of presentation were considered.

In the second experiment, attempts were made to improve upon conditions of the pilot study. Electronic equipment was modified to reduce probable artifacts, and achievement tests were examined for level of difficulty, reliability, and discriminative power. A sample of students not involved in the experiments was studied as a control. Data gathered from the control students were compared with those gathered from students in the actual experiment. This comparison made possible an analysis of the tests used, and a measurement of the equivalence of the tests in the control and experimental situations.

Difficulties in using the emotional response mechanism make it impossible to present pilot data gathered with that instrument.

CHAPTER FOUR: THE SECOND EXPERIMENT

REVIEW OF THE AIMS OF THE STUDY

The main purpose of our study was to compare the influence of liked and disliked teachers on student learning. We wanted data that would give us some idea on how to answer the following questions: (a) When students like the teacher, do they learn more than when they dislike the teacher? (b) Do the ratings that students give teachers relate to a measure of physiological arousal? (c) If a teacher is liked by students, can he stimulate more learning in them when he delivers the subject matter in person than when he is seen on film or heard on tape? (d) If a teacher is disliked by students, can he stimulate more learning when he is not seen—when he presents subject matter by tape recording—than when he is seen on film or in person? (e) Do students learn more when they are stirred up or aroused than when they are somewhat serene?

We used the term "emotional noise" to mean the effect of extra stimulation on the student. "Extra" or "extraneous" stimuli simply meant those things that the student could see and hear that were not central to the information to be learned.

The pilot experiment was an exploratory effort to help us find the faults that were almost certain to exist in any first attempt. The following points give an outline of changes that we made in our procedures as a result of the weaknesses found in the pilot study.

CHANGES MADE AS A RESULT OF THE PILOT STUDY

Changes in the Electronic Equipment

The electronic unit which was used to receive the weak GSR impulses prior to their amplification was completely re-wired to produce more reliable measures. Our electronic engineer (a graduate of the University of Michigan's School of Engineering) built an entirely new input device that greatly reduced the need of readjustments during data collection. Apparatus used in the first experiment made it difficult to be certain that the pen tracings moving at different distances from the central position were reflecting equal intervals for the same amount of deviation traced. This is the problem of establishing linearity of measurement. Tests of the new equipment indicated that equal resistance changes traced on the paper tape at different distances from the central position could be considered essentially equal. That change, of course, enormously increased the reliability of our GSR data.

Assessment of Achievement Tests

All achievement tests were examined for reliability by the Kuder-Richardson technique. Results of the computer analysis of both true-false (factual) items and multiple-choice (inference) items are presented in the following table.

TABLE 12

RELIABILITY COEFFICIENTS OF ACHIEVEMENT TESTS
COMPUTED BY THE KUDER-RICHARDSON TECHNIQUE

Type of Test	Test					
	1	2	3	4	5	6
True-false	.85	.77	.81	.78	.89	.87
Multiple-choice (Inference)	.96	.97	.97	.97	.97	.96

The above table shows high reliability for all inference or multiple-choice tests. The true-false or factual tests, however, hover on the outskirts of acceptable reliability. The second, third, and fourth true-false tests suggest that their scores be interpreted with caution.

Changes in the Sample of Students

A total of 48 ninth graders were selected from The University High School, Ann Arbor, Michigan. Selection was not made on the basis of randomization, because the attempt to establish two definite groups required a systematic choosing of students. We wanted both rapid and slow learners in our sample. Students were picked from an inspection of test profiles maintained in the school. A number of growth and development measures of the students had been taken over the time they had attended the school. The principal measures used to determine rapid and slow learners were: reading comprehension, intelligence, plus the curve plottings for organismic age (a composite of both psychological and physical measures, that included reading comprehension and mental age.) If the organismic curve had maintained itself above the average level of development during the history of testing and if both mental age and reading age were well above average, we regarded students with such records as "rapid" learners. If the plottings and measures were consistently below the hypothetical averages for normal development, students having such records were deemed "slow" learners. We soon discovered that in the University High School ninth grade only a relatively small number of students were slow learners. Consequently, we were unable to find as many slow learners as rapid ones.

Because of the time required to modify the electronic equipment in preparation for the second experiment, the period for data collection came closer to final examination time than was originally planned. The proximity to final exams caused irregularities in meeting the schedule. We were able, however, to get complete records on 21 students, that is, to collect all measures for six presentations. The other students completed less than six sessions.

Changes in the Subject Matter

After considerable effort to find or build messages suitable for our purposes, we discovered that portions of the reading comprehension test in the Iowa Tests of Educational Development were the most promising. One advantage connected with the borrowed passages was that carefully made test items were available and those items were judged to measure more than factual recall. We noted that the student not only had to understand the material to get a good score, but he also had to perform the kind of reasoning that was appropriate for our inference tests. So we used both the reading passages and the tests. In addition, we made 15 true-false items on each of the six units for testing of factual knowledge.

The subject matter was altered slightly to make all messages approximately 500 words in length. We tested the emotional appeal of the reading passages on a sample of 20 students in advance of data collection. The purpose of that preliminary run was twofold: to get data on the reliability of the tests, and to find out how students rated the content on semantic differential scales. We found that the tests were somewhat reliable, as already shown in Table 12. We also found that students rated the subject matter, all six units, very near to the neutral position; that is, their ratings showed that they had no strong feelings about the topics, as measured by mean scores.

Selection of Teachers

The last major change was in the selection of actors to assume the roles of positive and negative teachers. We used a sample of students in the ninth grade, from the University High School, to listen to each of eight actors, who assumed the teacher roles. They rated each one separately on the six semantic differential scales. Results indicated that the positive teacher used before should be retained; but that the negative teacher should be replaced by a contestant who received lower ratings. The differences between the average ratings of the two teachers were highly significant, showing a large difference in semantic space. The positive teacher was rated 31 in comparison with a possible maximum score of 42. The negative teacher was rated slightly under 15 in comparison to a possible minimum of 6.

Except for the above changes, the second experiment was the same as the pilot experiment.

SUMMARY OF THE PROCEDURES

In their order of occurrence the main procedures before and during the experiment were:

(a) Forty-eight ninth graders were selected from the entire ninth grade class of the University High School, Ann Arbor, Michigan, in the spring semester of 1965, on the basis of their growth and development records. We tried to choose as many slow learners as rapid ones, but we found that the number of slow learners in the population was considerably less than the number of rapid learners. The exact population was limited to those students in the ninth grade attending the University High School, Ann Arbor, Michigan, during the spring semester of 1965.

(b) Six independent units of subject matter were selected from the reading comprehension tests of the Iowa Tests of Educational Development. Test items on the reading passages were also borrowed from the same source.

(c) All messages or units were made comparable in length, totaling approximately 500 words each.

(d) A sample of ninth graders, not used in the teaching sessions, was selected as a jury to rate the performance of eight actors, who assumed the desired teaching roles. Two actors were selected from the eight: one as the positive teacher and another as the negative teacher. None of the students were aware that the actors were not teachers. As noted before, we chose experienced actors because of their convincing performances.

(e) All tests were examined for reliability and changes were made in items until we had a collection of acceptable tests.

(f) All tests were carefully checked for content validity. Items were discarded if judges, experienced in test making, disagreed about the appropriateness of them in relation to the subject matter. Also, items were discarded if at least one judge out of three deemed the item ambiguous.

(g) All electronic recording equipment was re-checked prior to data-collection. The major modification of the equipment has been noted earlier in the chapter.

(h) Students were scheduled for experimental sessions during times convenient to their regular activities. Because the time of data-collection came just before final examinations, we were able to get only 21 records of students who sat through all six presentations. For purposes of analysis, we considered the sample of 21 sufficient.

(i) The order in which students received the messages was randomized for each group. Four students at a time occupied the experimental classroom. Randomization of sequence was considered important to minimize possible bias that could have resulted from a fixed sequence.

(j) All students who completed the series were presented six separate messages, three that were delivered by the positive (liked) teacher and three by the negative (disliked) teacher. Each teacher delivered messages by tape, film, and in person (live).

(k) During the presentation of each message, GSR arousal patterns were recorded for each student.

(l) Immediately after each message was delivered, students were tested on the material by 15 true-false items, covering the facts, and by multiple-choice items, covering inferences based on the subject matter. Inference items were borrowed from the tests that accompanied the reading passages in the *Iowa Tests of Educational Development*.

(m) Students in the ninth grade of the University High School had not been examined previously by the *Iowa Tests of Educational Development*.

(n) One student during each session was asked to use the emotional response device for the recording of positive and negative feelings during the learning session.

RESULTS IN RELATION TO THE HYPOTHESES

To remind the reader of the hypotheses, they are repeated below:

(a) *Arousal is a function of the mode of presentation*. In order to support this hypothesis, it was necessary to observe that live presentations stimulated the most arousal, and that taped deliveries stimulated the least arousal. Arousal was measured by the number of galvanic skin responses (GSRs) and by conscious ratings of teachers and content on semantic differential scales. We thought that the greatest number of active stimuli would occur during live presentations, and the least during taped deliveries, with film in between and closer to the live presentations.

(b) *Under the negative teacher, achievement scores will decrease from tape to film and from film to live.* We expected to find that the disliked teacher would produce more emotional noise when delivering messages in person than by tape because of the added visible cues stimulating negative feelings. We also thought that learning from film would result in scores between those of tape and live presentation. Consequently, we expected that the amount of learning from highest to lowest would be in the order of: tape, film, and live.

(c) *Under the positive teacher, achievement scores will follow the same pattern as under the negative teacher, but the gradient will be more gentle and scores for the separate tests will be higher than under the negative teacher*. The reasons we had for adopting the above hypothesis were not based on comprehensive knowledge, and there was a lack of information from related studies. We thought that the liked teacher would produce some emotional noise, but less than the negative teacher. Also, we expected that some of the visual cues supplied by the positive teacher would facilitate learning and that the final result would be a canceling effect between positive cues and noise, with some drop in learning when compared with taped deliveries.

(d) Ratings of the same teacher will change across modes of delivery. The positive teacher will be liked most when presenting subject matter in person. And his positive effect will drop when presenting messages on film, and his lowest ratings, although above neutral, will occur during the taped mode of presentation. We predicted the opposite results for the negative teacher: least liked when seen in person, and rated highest (although still below neutral) when delivering messages by tape. His ratings on film would be close to the live presentations, but not so extreme in dislike.

Data for judging the validity of our first prediction are presented in Table 13 below.

TABLE 13

OBTAINED MEANS OF GSR FREQUENCIES UNDER THE VARIOUS TREATMENTS

Mode	Positive Teacher	Negative Teacher
Live	12.2	18.7
Film	10.4	22.8
Tape	13.2	19.3

We said that arousal, measured by GSR frequencies, was expected to produce a gradient, with the most arousal during the live presentation, and the least during the taped. Table 13 shows that the hypothesis was not supported by data obtained under either the positive or the negative teacher. Arousal was greatest during negative film and least under positive film. One corner of our expectation, however, was upheld; namely, that the negative teacher would produce greater arousal than the positive teacher. Arousal for the negative presentations was 60.8 for the total of the three means, while total arousal under the positive deliveries was 35.8. Aggregate averages were as follows:

Positive deliveries: 11.7 GSR arousals
Negative deliveries: 20.27 GSR arousals.

While our predicted gradient was not at all supported by the data, we can say that negative presentations produced nearly twice as many GSR arousals as the positive presentations. So far as these results are concerned, the negative teacher stirred up students much more than the positive teacher did. Table

14 shows the average ratings applied to both positive and negative teachers for the different modes of presentation.

TABLE 14

MEAN RATINGS OF TEACHERS ACROSS MODES ON SEMANTIC DIFFERENTIAL SCALES

Mode	Positive Teacher	Negative Teacher
Live	31.8	15.6
Film	33.2	12.4
Tape	29.4	14.6

The above table indicates that the positive teacher was rated well above the neutral value of 24, and that emotional arousal was greatest under the filmed presentation. The negative teacher produced ratings distinctly below the neutral value of 24. The strongest negative assessment was of the filmed delivery, that is, students rated it lowest. Using the number 24 as the neutral rating of the composite score from all six rating scales, the following shows how much the scores deviated from the neutral position.

Positive Live -	7.8	Negative Live -	8.4
Positive Film -	9.2	Negative Film -	11.6
Positive Tape -	5.6	Negative Tape -	9.4
	22.6		29.4

Because deviation scores were greater under the negative presentations than under the positive ones, it is reasonable to say that the negative teacher produced greater conscious arousal than the positive teacher. The means of these total deviations are 7.5 for the positive teacher, and 9.3 for the negative teacher. Since the two average ratings are in opposite directions from the neutral position, their distance is 16.5 points. We can say, therefore, that students felt much differently toward the two teachers.

Taking the results of both physiological and conscious arousal together, we find that the negative teacher produced more total arousal, than the positive teacher.

One common problem in using rating scales in research stems from the difficulty of knowing how truthful the ratings are: in filling out the blanks people often put down the ratings they think the researcher wants. Other people, however, may deliberately show ratings that they think will displease the researcher. Unless one has some other means for measuring the same thing, or a closely related thing, for comparison, the interpretation of ratings must rest mostly on faith and guesswork. The above measures show the kind of general agreement that tends to increase our belief in the truthfulness of the ratings.

The second hypothesis stated that achievement will vary under the negative teacher according to the means used for presenting the subject matter. We expected the highest achievement when messages were given on tape, because we reasoned that it would produce the least negative emotional noise. We also thought that, when the negative teacher delivered messages in person, the emotional noise would be highest and learning would be at its lowest. Film presentations were expected to yield results in between tape and live presentations, but closer to live than to tape. In short, we expected a gradient of achievement from highest to lowest in the order of: tape, film, and live.

Table 15 shows averages of the test scores from both the actual and the inference learning situations, as they were obtained under negative instruction.

TABLE 15

MEANS AND STANDARD DEVIATIONS OF ACHIEVEMENT SCORES
UNDER NEGATIVE CONDITIONS BY MODES OF PRESENTATION

4

Test	Live	Film	Tape
Factual			
Mean	7.43	7.90	2.19
s.d.	4.05	3.60	3.51
Inference			
Mean	13.09	13.86	9.00
s.d.	8.38	7.92	8.40
TOTAL	20.52	21.76	11.19

The standard deviations are large because tests were scored by a penalty formula that increased the spread of scores from negative values to positive values. Factual knowledge was measured by true-false tests, which were scored by the number of right answers minus the number of wrong answers. The inference tests were compased entirely of multiple choice items, each having four alternatives. Students were instructed to read each item carefully and to circle those alternatives that were incorrect and to be sure to leave the correct answer uncircled. The score for each multiple-choice test was computed by giving one point for each incorrect alternative that was circled and a negative three points for each correct answer circled. Such a scoring system is intended to provide more utility per item than the traditional method of scoring. It produces wide variations because the lowest possible score is -3N, where N is number of items, and the maximum possible score is +3N. For example, a test with 10 items of four alternatives each can yield a possible lowest score of -30 and a maximum score of +30. Students were told to leave blank those answers which they were not reasonably sure were incorrect. After a little preliminary practice in responding to items in that fashion, the technique offers little or no problems.

Table 15 shows that students learned best from the negative teacher when the subject matter was presented on film, and least when it was presented on tape. Again our prediction did not hold. The ranking order of test scores was the same for both factual and inference learning, suggesting that the mode of delivery had the same effect on both kinds of test behavior.

Table 16 presents results of learning under the positive teacher and in the same form as Table 15.

TABLE 16

MEANS AND STANDARD DEVIATIONS OF ACHIEVEMENT SCORES
UNDER POSITIVE CONDITIONS BY MODE OF PRESENTATION

Test	Live	Film	Tape
Factual			
Mean	6.28	5.61	6.43
s.d.	4.59	4.10	4.00
Inference			
Mean	10.60	10.55	12.96
s.d.	7.57	4.60	7.83
TOTAL	16.88	16.16	19.39

For factual scores, the gradient was almost reversed in relation to the predicted gradient. Results of the inference tests under the positive teacher showed that tape produced the highest scores, followed by live and film, in that order. When the means for both the factual and the inference tests are summed, the gradient descends from tape to live and from live to film, quite different from our prediction.

Table 17 presents the results of tests of difference between the means of factual achievement for the various modes of delivery and for both positive and negative teachers. Because the same students were involved in all six learning situations, the t-ratios were computed for correlated means. The table shows that 6 of the 15 t-tests failed to estimate significant differences. But 9 of the 15 did produce t-ratios significant to the .05 level or better. Levels of achievement, therefore, were not uniform across the conditions.

TABLE 17

MEAN DIFFERENCES BETWEEN FACTUAL TEST SCORES, SHOWING t-RATIOS AND SIGNIFICANCE LEVELS

Sign and Mode	Mean Difference	Standard Error of Mean Difference	t-ratio	d.f.	P
+L vs. -L	1.15(-L)	.53	2.17	20	.05
+L vs. +F	.67(+L)	.71	.09	20	NS
+L vs. -F	1.62(-F)	.70	2.31	20	.05
+L vs. +T	.15(+T)	.71	.02	20	NS
+L vs. -T	4.09(+L)	.87	4.70	20	.01
-L vs. +F	1.82(-L)	.80	2.28	20	.05
-L vs. -F	.47(-F)	.60	.78	20	NS
-L vs. +T	1.00(-L)	.61	1.64	20	NS
-L vs. -T	5.24(-L)	.86	6.09	20	.01
+F vs. -F	2.29(-F)	.54	4.24	20	.01
+F vs. +T	.82(+T)	.85	.96	20	NS
+F vs. -T	3.42(+F)	.80	4.27	20	.01
-F vs. +T	1.47(-F)	.75	1.96	20	NS
-F vs. -T	5.71(-F)	.82	6.96	20	.01
+T vs. -T	4.24(+T)	.67	6.32	20	.01

Legend:
+L is positive live
-L is negative live
+F is positive film
-F is negative film
+T is positive tape
-T is negative tape

Note: The symbol in parentheses just after number in the column headed "Mean Difference" indicates the condition that produced the higher mean.

Table 18 shows the results of mean differences between tests of inference, giving the t-ratios with levels of significance. Seven of the 15 comparisons produced t-ratios which estimate that the differences were stable and would occur with similar samples 95 or more times in 100 experiments. Eight of 15 differences were not significant.

For both factual and inference tests, 17 of the 30 comparisons produced significant t-ratios, suggesting that something beyond just chance variation was at work.

The experimental design and procedures did not allow us to quantify with precision the amount of influence of each of the independent variables. But further inspection of the data in a later section of this chapter is intended to narrow the possibilities of influence among the potentially contributing factors.

The third hypothesis, which predicted a similar gradient of mean scores under the positive deliveries, was not supported, as is indicated by Table 16.

TABLE 18

MEAN DIFFERENCES BETWEEN INFERENCE TEST SCORES
WITH THE t-RATIO VALUES AND SIGNIFICANCE LEVELS

Sign and Mode	Mean Difference	Standard Error of Mean Difference	t-ratio	d.f.	P
+L vs. -L	2.49(-L)	1.20	2.07	20	.05
+L vs. +F	.05(+L)	1.44	.00	20	NS
+L vs. -F	3.26(-F)	1.41	2.31	20	.05
+L vs. +T	2.36(+T)	1.40	1.69	20	NS
+L vs. -T	1.60(+L)	1.43	1.12	20	NS
-L vs. +F	2.54(-L)	1.60	1.59	20	NS
-L vs. -F	.77(-F)	1.77	.43	20	NS
-L vs. +T	.13(-L)	1.68	.07	20	NS
-L vs. -T	4.09(-L)	1.76	2.32	20	.05
+F vs. -F	3.31(-F)	1.49	2.22	20	.05
+F vs. +T	2.41(+T)	.81	2.97	20	.01
+F vs. -T	1.55(+F)	1.34	1.15	20	NS
-F vs. +T	.90(-F)	1.19	.75	20	NS
-F vs. -T	4.86(-F)	1.65	2.94	20	.01
+T vs. -T	3.96(+T)	1.32	3.00	20	.01

Legend:
+L is positive live
-L is negative live
+F is positive film
-F is negative film
+T is positive tape
-T is negative tape

Note: The symbol in parentheses just after the number in the column "Mean Difference" indicates the condition that produced the higher mean.

The fourth hypothesis suggests that student ratings of each teacher will vary across modes of delivery; that is, that these ratings will be significantly different between live and film presentations and between film and tape. The ratings were expected to increase for the positive teacher from tape to film and from film to live. It was also expected that the ratings would decrease for the negative teacher over the same order of presentations.

Table 14 shows that students rated the positive teacher highest for the filmed presentation and lowest for the taped delivery, whereas our prediction was that the live presentation would be highest and tape lowest. A more convenient form of comparison between expected and actual ratings of the positive teacher for the different deliveries, from highest rating to lowest, is:

Expected Ratings	Actual Ratings
live	film
film	live
tape	tape

For the negative teacher, a similar comparison between expected and actual ratings from high to low is:

Expected Ratings	Actual Ratings
tape	live
film	tape
live	film

In both cases, the expected order differed from the actual order. The filmed deliveries produced the greatest liking for the positive teacher and greatest disliking for the negative teacher, indicating that film stimulated more conscious arousal of student feelings than did the live condition. It therefore seems that film may be an effective medium for arousing the emotions of students both for and against a teacher. Perhaps the most effective use of films may be found in influencing the immediate motivation of students.

Although students ranked the same teacher differently under the different modes of delivery, it was found that only two of the six statistical analyses showed the differences to be significant. The sign-rank test, as developed by Wilcoxon, was applied to the ratings given to the same teacher. The following list shows that only one difference was significant for the positive teacher:

Live vs. film (f) — not significant
Live vs. tape (l) — not significant
Film vs. tape (f) — significant at .01 P.

The letter in parentheses, in each case, indicates the mode that received the higher rating. Only the ratings of the positive teacher for tape and film revealed a significant difference. The teacher was rated higher for film than for tape.

A comparable list of results for the negative teacher is:

Live vs. film (l) — significant at .01 P.
Live vs. tape (l) — not significant
Film vs. tape (t) — not significant

Students ranked the negative teacher higher when he delivered the message in person than on film.

Our fourth hypothesis, which predicted that the same teacher would be rated differently across means of presentation, was not confirmed. Significant differences were in the minority by two to one, when compared with data in which there were no significant differences. In general, it seems that the same teacher will be rated about the same by students, regardless of whether he conveys his message in person, on film, or on tape.

Summary of Results in Relation to the Predictions

In general, none of the hypotheses was well supported by the evidence.

(a) Arousal is a function of the mode of presentation. This prediction was unsupported.

(b) Hypotheses two and three indicated that a gradient for both the positive and negative learning sessions would produce similar patterns of achievement. Taped deliveries were expected to result in highest achievement because of the presumed low degree of emotional noise, and live presentations were expected to yield the lowest achievement. These predictions were not supported.

(c) The fourth hypothesis predicted that the same teacher would receive different ratings under different modes of presentation. In general, this hypothesis was not upheld by the evidence.

FURTHER RESULTS

Rather frequently, the most valuable results of experiments are not directly related to the hypotheses posed at the beginning. The following series of additional findings merits special mention.

1. The teacher was the most influential of all factors in producing arousal. The teacher far eclipsed the mode of presentation in producing arousal as supported by the fact that all tests of significance between positive and negative teachers within each mode of presentation were well beyond the .01 level of probability. This was true of arousal measured by GSRs and by semantic differential ratings.

2. The disliked teacher stimulated higher achievement than the liked teacher when messages were presented in person. Since the ratings between the positive and negative teacher were highly significant, with the positive teacher having much higher ratings, it is meaningful to examine their relative effects on learning.

(a) The factual test results indicated that the negative (disliked) teacher stimulated a greater amount of learning, significant at the .05 level of probability. Analysis was performed by the t-ratio, using the formula for correlated means because the same students were involved in all sessions.

(b) The inference test results revealed that the negative teacher stimulated more learning than the positive teacher. Results of this test favored the disliked teacher even more than those of the factual test. The t-ratio significant beyond the .01 level of probability.

3. The disliked teacher stimulated higher achievement than the liked teacher when messages were presented on film. That was true for both factual and inference tests. Comparisons made by the t-test technique were almost identical with the live presentations. The same levels of significance were found at the .05 level of probability for the factual tests, and at the .01 level of probability for the inference tests.

4. The liked teacher stimulated higher achievement than the disliked teacher when messages were presented on tape. This finding held true for both factual and inference tests, both differences being significant at the .01 level of probability.

5. For each of the six deliveries, the amount that the student learned had no relationship to the strength of his rating of the teacher. For example, students who rated the liked teacher at the very maximum learned no more than those who rated the teacher below the neutral point. The finding suggests that the scores students assign to teachers during course evaluation sessions cannot be interpreted in the common sense way, by assuming that those who rate the teacher highest have achieved the most. There are those who already concede this point, but would presist in maintaining that the ratings are indicative of how much the student is led to accept the subject matter. The importance of this opinion can be seen in the fact that if a student dislikes a particular subject matter, he is likely to choose another field of specialization.

The facts are indeed not sufficient to support the claim just stated, and besides, it seems likely that the matter is actually more complicated. It is doubtful, for example, that instructors in schools of medicine make particular efforts to be liked by students. Yet, the rate of dropouts due to disenchantment with medicine may be even less than in teacher training, where professors appear to be highly concerned about the feeling tone of students. Of course, this does not mean that a negative attitude towards both teacher and subject matter is to be preferred to a positive attitude, but it does suggest that the common-sense assumption should be carefully examined before it is accepted as a policy for assessing teacher effectiveness.

6. When students were taught by modes that combined both visual and auditory stimuli (live and film), the more they were stirred up or aroused, the more they learned. Table 19 shows results of combined GSR averages for both film and live presentations by the positive teacher and by the negative one. It also indicates total amount of learning under each teacher. The table shows that the negative teacher stimulated more learning and produced more GSRs than the positive teacher.

7. When students were given subject matter only by auditory means (sound tape), the more they were aroused the less they learned. Table 20 shows results in the same form as Table 19. The relation between arousal and learning is reversed in Table 20 as compared with Table 19. The obvious question is: Why? It is possible that, when a message is carried only by audible means, a single unit, or bit, of noise may be more damaging to effective communication than when

TABLE 19

AMOUNT OF PHYSIOLOGICAL AROUSAL (GSRs) AND LEARNING UNDER PRESENTATIONS USING BOTH AUDITORY AND VISUAL STIMULATION (LIVE AND FILM). SCORES SHOWN SEPARATELY FOR EACH TEACHER

Teacher	Combined GSRs Under Live and Film	Combined Learning Under Live and Film
Positive	22.6	33.04
Negative	41.5	42.28

Each number represents the sum of the two means involved.

TABLE 20

AMOUNT OF PHYSIOLOGICAL AROUSAL (GSRs) AND LEARNING UNDER PRESENTATIONS LIMITED TO AUDIO STIMULATION (TAPE). SCORES SHOWN SEPARATELY FOR EACH TEACHER

Teacher	No. of GSRs Under Tape	Learning Under Tape
Positive	13.2	19.39
Negative	19.3	11.19

Each number represents the sum of the two means involved.

both visible and audible stimuli are involved. In re-running the films, we noted that both teachers used gestures and facial expressions to emphasize certain points. Consequently, we cannot say that all the extra visible stimuli amounted to only noise. Our concept of "emotional noise" was not originally based on the recognition that many of the visual cues of the teacher may facilitate rather than just inhibit learning. Also, our findings in the above tables may suggest the operation of certain cultural influences, which will be given later, in the "Discussion" section.

8. Students who operated the emotional response meter during learning sessions produced tracings that were predominately negative under the negative teacher and predominately positive under the positive teacher. Results agreed in direction with the semantic differential ratings. Conscious feelings that students have during learning sessions apparently agree with ratings about the session completed soon thereafter.

9. Differences in measures on the basis of sex were not sufficiently significant to merit seperate presentation.

10. Students who had been classed as slow learners prior to the experiment rated teachers at the extreme ends of the rating scales. This finding is based on only four slow learners, all that completed the experiment. Because of the small number of slow learners, previous analysis has not stressed differences between them and rapid learners.

11. Measures of physiological arousal (GSRs) taken on slow learners were highly erratic in comparison with measures of rapid learners. There was a much less consistent pattern provided by GSR frequencies of slow learners under the different conditions of learning than provided by rapid learners. Further investigation of the phenomenon is needed using larger samples than our study provided.

12. In general, the more students rated the subject matter above the ratings given to the teacher, the greater the learning. Although it is possible that this finding could be an artifact produced by the conditions of arousal, it merits recognition as a basis for future study.

DISCUSSION

Research results often fail to support our common-sense opinions. It is likely that many, if not most, of our current opinions about how to improve education will not be confirmed by future research. Because the need for education continues to rise as social change becomes more rapid, the urgency for educational research increases accordingly.

Most of us perhaps have believed for a long time that emotion plays an important part in human learning. But we have hardly made a beginning in understanding the complex interplay between feeling and knowing. Our present study is just a tiny part of the research that is needed before we can say with much assurance just how individual students should be stimulated to help them get the most from education.

At the beginning of our experiments, we assumed that more stimuli would bear upon the student when he listened to the teacher in person than when he simply heard him on tape. Consequently, we expected more signs of arousal (GSRs) under the live condition than under the taped delivery. But the facts did not agree with our thinking. More arousal was produced when students listened to tapes than when they both saw and heard the teacher in person. Although these differences between tape and live delivery under the same teacher, were not significant, the results force us to re-think our original position. The obvious implication of our findings is that the size of a population of external stimuli bearing upon a student at a given time has little or no connection with the amount of his internal arousal. The quality and intensity of stimulation are probably more important than simply the number of stimuli. We are now inclined to think that the amount of attention and concentration required by the student to grasp a message is perhaps the important factor in producing arousal. The idea could probably be easily tested by presenting two matched groups of subjects the same message on tape with only one change, the volume or loudness of the message. The first tape run would be given at the normal level of loudness while the second would be greatly reduced to the point that the student would have to strain to hear it. In the second delivery, students would have to give their utmost attention and concentration to the presentation in order to follow the message. We would expect more arousal when the volume is reduced. Would there be any difference in the amount of learning under those two conditions? Common sense would suggest that more learning would occur under the normal level of loudness. But we are inclined to expect that either no difference would occur, or if a difference did occur, that more learning would be found when the volume is reduced than when it is at a normal level. The suggestion points up one of many studies that are needed before we can describe the complex procedures that make up teaching and learning.

More conscious arousal was produced by the films than by the other two means of delivery. This was true for both the positive and the negative teacher. A moving picture projected on a screen normally produces a sharp contrast with the immediate surroundings, which are partially blacked out by the dimmed lighting. Visual stimuli are thus more effectively focused, or concentrated, than is the case of a live presentation in fully lighted surroundings. We not only found the greatest conscious arousal for filmed deliveries for each type of teacher, but we also found the best learning. Tables 13, 14, 15, and 16 show that the positive teacher was more effective on film than he was on tape or in person. And the same was true for the negative teacher. If we can assume that our results would occur if the experiments were repeated for different samples of students and subject matter, perhaps we should exploit the use of motion pictures more than we do in teaching.

The idea that we called "emotional noise" is meaningful, but our original conception of what produces it now seems somewhat primitive and inadequate. We were correct in thinking that the negative teacher would produce more arousal than the positive teacher. But we were evidently wrong in thinking that most of such arousal could be taken as emotional noise, which would interfere with learning. For both filmed and live deliveries, arousal correlated highly with amount of learning, the more arousal the more learning. Obviously, we cannot say that the arousal was mostly emotional noise that served to diminish learning. The significant fact is that students showed a strong dislike for the negative teacher in comparison with the positive one. But despite that difference, which was highly significant, students learned more from the negative teacher under both film and live conditions than from the positive teacher! Even if the disliked teacher did produce more emotional noise than the liked teacher, he must also have produced more cues for learning that more than compensated for the noise. Effective cueing is perhaps the most important part of the teaching art. As mentioned before, re-runs of the films of both positive and negative teachers suggest, on the basis of ex post facto consideration, that the negative teacher was quite effective in producing facilitating cues. His movements were very appropriate for emphasizing certain points in his message.

Results showed that although the students disliked the negative teacher they did attend to his presentation. We had originally guessed that dislike for the teacher would predispose to inattention, but again our preconception was unsupported by the facts.

Attention is necessary for learning, and the teacher who succeeds in capturing and holding it has a great advantage. Positive or negative feelings by the student toward the teacher apparently do not guarantee attention so much as other factors, such as cueing. Even a teacher who is disliked, but who produces effective cueing, can, according to our results, stimulate greater learning than some teachers who are distinctly liked.

We noted that the negative teacher was much more authoritarian than the positive teacher. He gave the impression that his words were not to be disputed, and that to do so would be an admission of ignorance rather than evidence of good thinking. Of course this impression is strictly subjective, and we cannot produce any measurable evidence for it, but as Cronbach and others have already pointed out, everything that an authoritarian teacher does is not necessarily detrimental to learning. In fact, some things that he does may facilitate effective learning. One such factor may be the ability to hold the attention of the students. The more evidence that accumulates upon such current issues in education as authoritarianism versus permissiveness, the less likely it seems that all the desirable characteristics are on the side of the currently preferred value. Research promises to modify our educational values rather than merely expressing them.

As we mentioned late in our discussion on the first experiment, the greater effectiveness of the negative teacher over the positive teacher on film and in person may be related to a traditional influence of teachers in our society. Judging from what many students say, they seem to prefer vacation time to school attendance. They like some things in formal education, but the weight of evidence suggests that their feelings towards classroom learning are slightly below the neutral point, or slightly negative. If that is true, then it is easy to understand why the threatening teacher is more effective in holding attention and getting certain tasks performed by students than the permissive teacher. We do not mean that threat is actually a better means of control than positive reinforcement, but we do mean to suggest that effective factors of teaching relate to the cultural traditions in a society. If American education carries the tradition of control through threat and punishment to a greater extent than control via positive reinforcement, it is reasonable to think that the likeable and permissive teacher will have to operate for some time before the advantages expected from his type of management will produce the expected payoff. Our experiments were too short in duration to test this hypothesis, but other studies have shown that the immediate results of changing from an authoritarian to a permissive type of management are some loss rather than any gain in learning. Field studies that range over considerable time spans are needed to test the implications of short term experiments.

In order to judge the effectiveness of a teacher, our experiments suggest that the ratings of the teacher alone may be less valuable than the differences between the teacher ratings and subject matter ratings. To get this kind of evidence, one uses comparable semantic differential scales for both the teacher and subject matter. Our findings show that for any given learning session, ratings of the teacher by students told us little or nothing, but when we looked at the total difference between ratings of the teacher and of the content we found a potentially useful measure. If, for example, a teacher is very much liked, as shown by high ratings, and if the subject matter falls far below teacher ratings, the level of achievement is likely to be low. The suggestion is that both teacher and content ratings ought to be fairly well balanced in order for a liked teacher to stimulate high achievement. But the picture is somewhat reversed for the disliked teacher. When the disliked teacher somehow succeeds in maintaining a reasonably good opinion about the subject matter, he can be rather effective. Our tentative principle may be stated as follows: A teacher liked by students can be effective if his personality does not overshadow the student's interest in the subject matter, and a disliked teacher can be effective in stimulating learning if dislike of him does not spill over too much in to the subject matter. Two popular teachers may be liked for different reasons. If the popularity of a teacher shows high transference to his subject matter, so that the positive feelings by students are reasonably well balanced between the teacher and the course, we suspect that such a teacher will stimulate optimal learning. On the other hand, a second teacher may be popular with students in a way that has little or nothing to do with transfer of positive feelings to the subject matter. In that case, the teacher will stimulate scant achievement. Of course, teachers can also be disliked for different reasons, but if the dislike fails to transfer to the subject matter and the students rate the content high, the teacher can be more effective in helping students achieve than some popular teachers.

Erratic reactions of slow learners measured both by GSRs and by the semantic differential scales are somewhat puzzling. In rating teachers and content, slow learners tend to be more apathetic toward the teacher when he is both seen and heard than when the subject matter is presented by sound tape. Also, slow learners, based on our scant evidence, were found to rate disliked teachers much higher than rapid learners while rating liked teachers much lower than rapid learners. If this finding held in larger samples, it could lead to valuable modifications in our expectations about slow learners.

In our study, slow learners behaved inconsistently in terms of physiological responses. Their GSR frequencies tended to be at the extremes of the distributions, either at the high end, reflecting a great deal of activity, or at the low end. We suspect that under given conditions, there is a certain range of GSR frequencies that correlate with optimal achievement, and perhaps two other segments that correlate with low achievement. We thought our data insufficient to make strong claims about the specific function relating GSR frequencies to learning, except for the tendency of high arousal and high achievement to occur together when students were taught by visual and audio stimuli together.

Our experiments carry some implications about the issue of theory-based research versus empirical studies, which are not based on well structured theories. We believe that this issue often results in fruitless debate, but we also think that the issue can be meaningful if it centers on the shortcomings resulting from rigid adherence to either side. Although our hypotheses were probably not based on the best available rationale, our study would have been worthless if we had simply tested the hypotheses and not looked beyond them. Although, because of limitations already noted, our additional findings cannot be used to support any broad generalizations, we still think that the chief value of our study lies precisely in the implications of these additional findings. It may be that empirical research has more justification than hypothesis-testing research when the results of related studies are not sufficient to develop a strong theoretical position. If so, the best thing for the researcher to do is to pose a series of questions that honestly reflect the present state of ignorance about the problem area, and not to theorize with an aim of self-confidence that he really does not have. We suspect that for a long time to come that there will be a need for more empirical research, and that only after such research will we have sufficient knowledge to develop a sound theory of teaching and learning.

The influence that teachers have on students is no doubt extremely complicated. We think that the terms "positive" and "negative" to express a student's feelings toward his teachers are too gross to be good as descriptive terms. Our results did not confirm the idea that, when a student simply likes a teacher, he learns better than when he dislikes a teacher. "Like" and "dislike" are blanket words that cover too many specific feelings. We found that effective learning often occurred when a student registered dislike for a teacher, and that poor learning could occur when the teacher was liked. Future studies should probably go into the factor analysis of the syndrome of the so-called positive and negative feelings. It will probably be worthwhile to attempt to discover the particular aspects of the teacher that stimulate emotional reactions and to determine what kinds of feelings tend to generalize from the teacher to the subject matter. If we could specify such things reasonably well, we would probably have the kind of knowledge that could be used with great effectiveness in teacher training and in classroom management.

One of the big problems in studies of feelings and emotions lies in this question: "How much can we generalize about conditions that stimulate emotion when we are using groups of persons?" Individual differences are so great and cover such a wide range that it seems necessary to make intensive studies of individuals before we can discover the value of research on groups. Although many case histories have been reported, they usually fail to show the influence of rather specific variables on behavior. We apparently need a new approach, something that could be termed "the single student experiment." So long as we could show changes in the student's behavior as a result of certain things that happen to him, we could develop a record of functional relations between his behavior and environmental changes. Development and investment in the single student experiment may lead us to the kind of information needed to make education far more effective than it is.

CHAPTER FIVE: SUMMARY AND CONCLUSIONS

SUMMARY

Do students learn more from teachers they like or from those they dislike? Are students stirred up more physiologically by a disliked teacher than by a liked one? When students are aroused or stirred up by the teacher do they learn more than when only slightly aroused? If a liked teacher delivers his subject matter by tape, on film, and in person, which means of delivery stimulates the most learning? Which of the three means for presenting subject matter is most effective when the teacher is disliked? It was our attempt to answer these questions that prompted our experiments.

We first set up a pilot experiment to explore the promise of the method that we thought would yield appropriate information. Among the available classes of physiological arousal, we choose the galvanic skin response because it had been used by psychologists for over 40 years and in many situations. The galvanic skin response (usually abbreviated as GSR) is a change in the electrical resistance of the skin and has long been one of the measures recorded by the lie detector. Our purpose was not to detect lies, but to record GSRs as evidence of physiological arousal. We assumed that the more GSRs found during a learning session the more a certain kind of activity occurred inside the skin. We also measured the size or amplitude of the GSRs as well as the time required for each one to reach its peak from the beginning. We called that amount of time the "rise time."

We also recorded the conscious feelings of students towards teacher and subject matter by presenting students with rating scales of the kind developed by Osgood, who named them "semantic differential" scales. (See Appendix A) Much research has been reported on the semantic differential, which is now believed to be the most appropriate measure of the feelings or emotional posture of a person towards any of a great variety of stimuli, including words, things, persons, and small and large organizations of various kinds. We believed, therefore, that our use of the semantic differential for estimating how students felt about teachers and subject matter was proper.

We selected from a pool of experienced actors the two which best assumed the roles of the positive and the negative teacher. The positive teacher was the one rated high on the semantic differential scales; that is, he was strongly liked by the students. The negative teacher was rated low; he was disliked by students. The choice of the best actor for each role was made by a sample of adults and high school students, before whom they acted, and who rated them along such dimensions as "likeable-annoying," "irritating-pleasing," "good-bad," "unfriendly-friendly," and the like. None of the students were aware, so far as we could determine, that the actors were not teachers.

We thought that the means of delivery used by the teacher for the subject matter would have some effect on the strength of student ratings, upon physiological arousal, and upon learning effectiveness as measured by achievement tests. The tests of amount learned (achievement) were of two kinds: a set covering factual information and a set that tested how well students could make logical inferences from the subject matter—a test requiring critical thinking.

We had both the positive and negative teacher present messages by tape recording, on film, and in person to the student audience. Our experimental classroom could accomodate only four students at a time because of space limitations. Each student sat at a small desk that was wired in such a way as to make it possible for ring electrodes to be easily connected to the first and third fingers of each student's non-writing hand. Changes in electrical resistance of the skin were picked up by an electronic device in an adjoining room. The electrical impulses were then fed to an amplifier and from there to a recorder, which traced lines corresponding to the changes in skin resistance. Our equipment allowed us to get permanent GSR records of each student during each of the six learning sessions.

The six learning sessions were labeled as: positive and negative live (teacher in person), positive and negative film, and positive and negative tape. The order of the presentations was randomized for each group of four students so that the possible effect of a fixed sequence would be avoided.

In the pilot study we used college students, mostly sophomores, enrolled in a psychology course at the University of Michigan in the fall of 1964. Results of the first experiment showed that our procedure and equipment needed certain improvements, as listed in the early part of the preceding chapter.

After the changes had been made, we conducted a second experiment, using ninth graders in the University High School, Ann Arbor, Michigan. Because the results of the second experiment were more reliable than those from the pilot experiment, the following list of results refers only to the second experiment. Before the results are summarized, a brief review of the hypotheses that guided the experiment must be given.

We stated four such hypotheses, which we drew from our reflections on the possible relations involving arousal, feelings, and achievement. They were:

(a) Arousal is a function of the mode of presentation. More arousal will be produced by the teacher in person than when he is heard on tape, because more stimuli will bear upon the student then when he receives the subject matter by hearing alone. Film will arouse students almost as much as live delivery, and more than tape.

(b) Under the negative teacher, achievement scores will be highest for learning from tape and least from live presentation. The gradient of achievement, involving only the negative teacher, will be in the following order from highest to lowest: tape, film, live. When students dislike a teacher to a marked degree, emotional noise will result. More emotional noise will result from live presentation than from tape because of the greater number of stimuli presumed to be active in the live situation. Dislike of the teacher will tend to cloud communication. Therefore, scores on tests will be relatively low when students confront the disliked teacher in person, and because the taped presentations will not press many extra stimuli on the student, over and above those required to deliver the message, achievement scores will be highest when the subject matter is carried by tape.

(c) Presentations by the positive teacher will result in a gradient of achievement similar to that obtained from presentation by the negative teacher. The only expected difference between the two gradients will be that the one produced by the positive teacher will be gentler than the other. The liked teacher will produce emotional noise that will lower scores on the inference tests because students will be prone to somewhat excessive and uncritical acceptance of the things given by the positive teacher. In short, the positive teacher will not be likely to provide a good mental set for critical thinking.

(d) And lastly, ratings by students of the same teacher will change according to the means used to deliver the subject matter. The liked teacher will receive the highest ratings when he presents the material in person. Ratings for film will be lower than for live presentation, and ratings for tape will be lowest. The reverse order will prevail for the disliked teacher: He will be least liked when presenting material in person, liked slightly better on film, and liked best on tape.

Results did not bear out any of our predictions significantly, but we did find that the data revealed a number of other findings that justified our efforts. The additional findings are:

(a) The disliked teacher produced far greater physiological arousal (number of GSRs) than the liked teacher. Although we expected this result, we did not specifically hypothesize it.

(b) When both teachers presented material in person, students learned significantly more from the negative teacher than from the positive one.

(c) When both teachers presented subject matter on film, students again learned more from the negative teacher.

(d) When messages were given via tape, students learned more from the liked teacher than from the disliked one.

(e) Measured differences according to sex were not sufficiently great to warrant special emphasis.

(f) Slow learners produced erratic patterns of physiological arousal; that is, their patterns of arousal were less consistent and systematic than those produced by rapid learners.

(g) Amplitudes and rise times of the GSR tracings were not found to be as useful as GSR frequencies.

(h) Slow learners rated teachers either very low or very high. In other words, their ratings were extreme in comparison with ratings made by rapid learners.

(i) Under the disliked teacher, those students who rated the subject matter relatively high obtained higher scores on achievement tests than those who rated the subject matter at the low end of the distribution.

(j) Under the liked teacher, those students who rated the subject matter near to the teacher rating received higher scores than those who rated the subject matter much differently than they rated the teacher.

(k) For any given presentation, there was no relationship between amount learned and ratings assigned to the teacher.

(l) The teacher was far more effective in producing arousal than the means used to deliver the subject matter.

CONCLUSIONS

Although it is not tenable to make broad generalizations from the results, the following conclusions are deemed possible when considered under the limitations of the study.

(a) If students have a low opinion of the subject matter, despite the effects of a positive teacher, achievement is likely to be medium to low. It seems that if students hold a low opinion of the subject matter, regardless of whether or not they like the teacher, their achievement will be considerably below their potential for learning.

(b) A high student rating of a teacher is no guarantee that the student also likes the subject matter.

(c) For the majority of students, a medium to fairly high level of physiological arousal appears to be an important correlation to high achievement. The teacher who fails to stir up the feelings of students during instruction is less likely to stimulate high achievement than the teacher who creates considerable arousal. That appears to hold whether the teacher is liked or disliked.

(d) When messages are presented by tape recording it seems to be important that the speaker's voice be pleasant and that the speaker maintain a continuous delivery. We found that the least amount of learning occurred under the negative teacher on tape. The quality of his voice was far less pleasing than the liked teacher's and he failed to speak continuously, as the positive teacher did, who generated much higher achievement by tape.

(e) Students can endure considerable negative stimulation by a teacher and still learn quite well when the teacher is both seen and heard. One possible reason for this is that students apparently depend upon visual cues to supplement the oral message. And if those cues are appropriate, most students will learn rather well despite their personal feelings toward the teacher.

RECOMMENDATIONS

Judging by what many students have to say about courses in education, it seems that the intellectual respect they accord to those courses is less than to those in subjects taught in most other university departments. If this is a reasonably valid assessment (and it needs a careful empirical test), then efforts to improve the prestige of courses in education may be a better investment than concern over the popularity of professors. The challenge provided by the problems in education is no doubt very great and will tax to the limit the best minds available. But if our courses do not reflect that challenge, and students consider them embarrassingly naive, our main concern should be a drastic revision of subject matter.

Research designed to measure teaching effectiveness should be so conceived that serious recognition is given to the traditions of education peculiar to the society and sub-culture under study. Predictions that are tested without due consideration for cultural variables are likely to result in information that is far less useful than when the influences of tradition are clearly recognized.

Because group experiments in education are extremely hard to design and carry out in a manner clearly distinguishing all variables, it seems highly advantageous to place increased emphasis upon the single student experiment. This type of experiment promises to offer ways of avoiding many of the difficulties usually found in group studies. If the problem of individual differences is as large as we think it is, it cannot be solved until we devise ways of securing information about the extent and importance of individuality. It seems that to get such information, we must rely heavily on single student experiments which have certain advantages over the familiar case study. In general, the advantages lie in the direction of the increased controls maintained in experiments, as compared with the usual kind of case study.

We suggest that instruments used to rate teachers be improved by adopting semantic differential scales. Separate ratings should be made by students on the content. After the two ratings have been recorded, the difference between them should be more useful than the two scores taken separately. If students are found to rate a teacher high while rating the subject matter low, the disparity between the two ratings will strongly favor the teacher. We believe that such disparity is more likely to be correlated with relatively low achievement than when subject matter is rated as high as the liked teacher. If the teacher is disliked by students and if the subject matter receives a high rating, the level of achievement should be reasonably good, but if both the teacher and subject matter are rated low, learning is apt to be considerably below what it should be. We believe that it would be worthwhile to use this recommendation as a hypothesis, and that several studies be completed before any substantial claim is made.

The phenomena of physiological arousal appear to be significantly related to certain processes of education. We recommend the kind of experimental setup that is capable of acquiring data both on arousal and on conscious feelings during the learning session. The research components should be capable of storing and analyzing such data very rapidly so that the time between data collection and analysis can be reduced. The computer appears to hold the most promise for rapid progress from data-gathering to data-analysis in the conduct of research into teaching and learning, given the vast quantities of diverse information generated by such research. We believe that a system which can perform this task can be used to improve teacher training by providing feedback to the supervising teacher and to the trainee at the time it is most needed.

Until we have more adequate descriptive information concerning the multiple relations of such variables as conscious feeling, arousal, various kinds of achievement, modes of delivery of information, and aspects of the teacher, we shall not be in the best position to produce a good theory of teaching. Such a theory is urgently needed to guide our research and to pull together into a meaningful whole the many separate findings on instruction. We believe that the most promising tools for describing classroom processes are found in computer-based systems that have the power to store data as they occur during the process of teaching and learning. Rapid analysis of such data, reduced so that results are easily understood, could be most promising for improving teacher training. The speedy means of feedback offered by a computer-based system should be valuable for both the supervising teacher and the trainee, working together to modify their management of the classroom situation in the light of the effects of their previous efforts. We also think that current attempts to set up computer-based systems of instruction are likely to yield slight returns until we know a lot more than we do now about the relationships between the many variables involved in classroom instruction. Once we have enough descriptive evidence and a good theory of teaching, we shall be able to use all the media of teaching and learning much more effectively than we do at present. Until that time comes, the cost of research into computer-based instruction is likely to be unnecessarily expensive.

APPENDIX A

SEMANTIC DIFFERENTIAL SCALES

How I Feel About the Subject Matter

valuable_____:_____:_____:_____:_____:_____:_____: worthless

irritating_____:_____:_____:_____:_____:_____:_____: pleasing

good_____:_____:_____:_____:_____:_____:_____: bad

uninspiring_____:_____:_____:_____:_____:_____:_____: inspiring

sound_____:_____:_____:_____:_____:_____:_____: unsound

boring_____:_____:_____:_____:_____:_____:_____: interesting

How I Feel About the Instructor

likable_____:_____:_____:_____:_____:_____:_____: annoying

negative_____:_____:_____:_____:_____:_____:_____: positive

superior_____:_____:_____:_____:_____:_____:_____: inferior

unfriendly_____:_____:_____:_____:_____:_____:_____: friendly

interesting_____:_____:_____:_____:_____:_____:_____: dull

silly_____:_____:_____:_____:_____:_____:_____: clever

APPENDIX B

INSTRUCTIONS READ TO EXPERIMENTAL SUBJECTS

This is an experiment in human learning. The purpose of the experiment is to examine the relationship between how well you can learn and recordings of your skin resistance. The electrodes on your fingers will pick up tiny changes in electrical conductance which will be recorded on graph paper in the next room. There is no electrical shock involved. The important thing for you to do is to listen very carefully to the material presented to you. Try to remember as many of the details of the message as you can. You will be tested immediately after each message, which will last about five minutes.

APPENDIX C

DIRECTIONS FOR TAKING THE MULTIPLE-CHOICE TESTS

1. Do not circle the answer you think is correct, instead

2. Circle each answer that you believe is false, being careful not to circle the correct one.

3. You will be scored one point for each incorrect answer that you circle.

4. If you circle the correct answer three points will be deducted.

5. Circle only those wrong alternatives about which you are rather certain are incorrect. Pure guesswork on your part is very likely to lower your score.

6. Remember, circle only those alternatives in each item that you are reasonably sure are incorrect. Leave all the rest unmarked.

APPENDIX D

IDENTIFICATION OF SUBJECT MATTER FOR THE SECOND EXPERIMENT

The following selections were taken from the Iowa Tests of Educational Development, those sections which contained passages for testing reading comprehension appropriate for ninth grade students. The units which are numbered below are the designations used only in the experiment.

For Unit 1. Test 5, Interpretation—Social Studies. Article titled "U.S. Population in 50 Years May be 200,000,000." Entire article was used.

For Unit 2. Test 5, Interpretation—Social Studies. Untitled article was used which treated collective bargaining, incidents in 1938-1940. Entire article was used, with some padding to increase length.

For Unit 3. Test 5, Interpretation—Social Studies. Untitled article was used discussing Cardinal Richelieu. Last two paragraphs were not used, except for the first sentence in the next-to-last. An additional sentence was added which read, "While leaving to the nobles nearly all their privileges and their wealth, he turned over the public business more and more to the middle-class officials."

For Unit 4. Test 5, Interpretation—Social Studies. Untitled article was used discussing economic possibilities in Latin America. Entire article was used, but some unessential words were deleted in order to reduce the length.

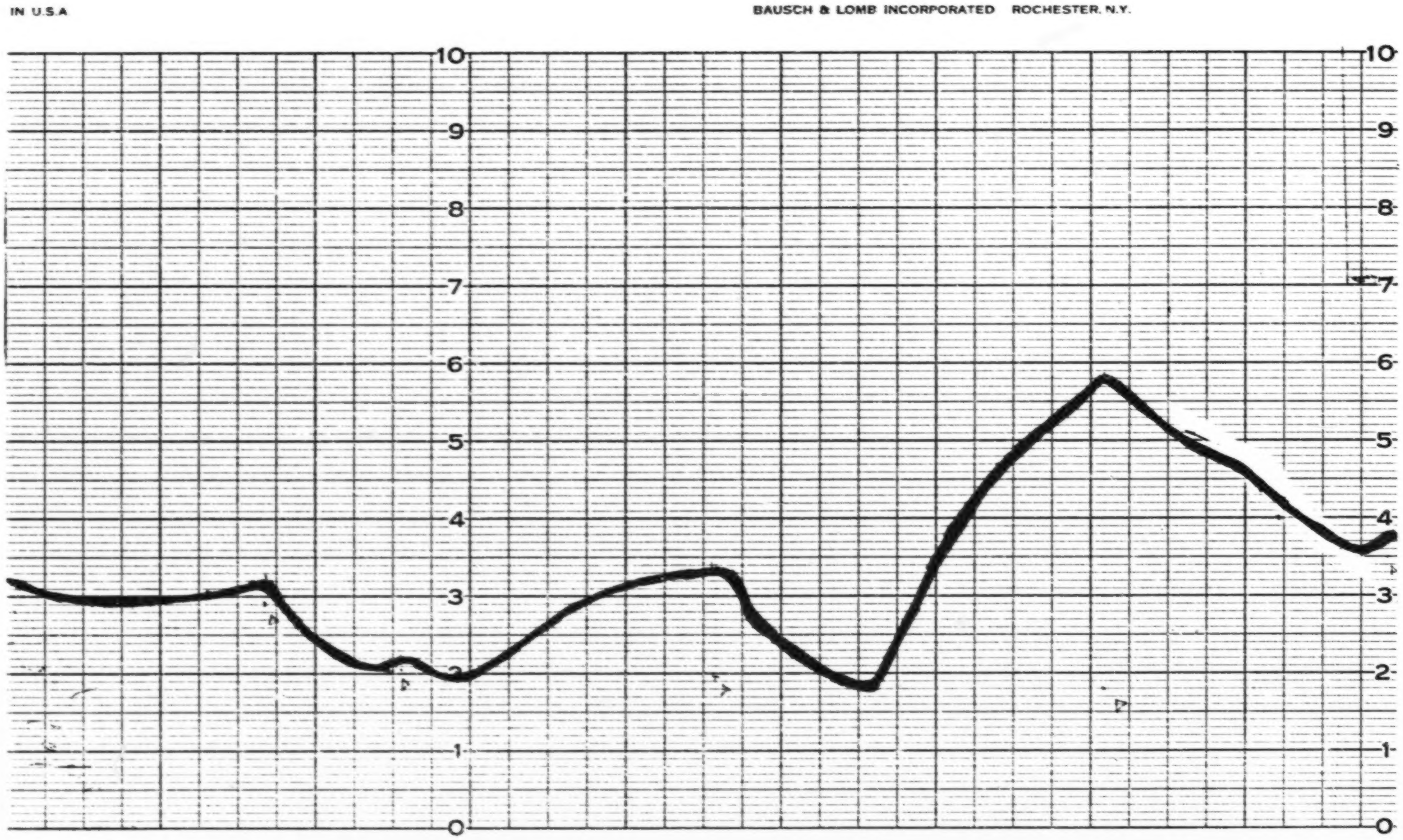
For Unit 5. Test 5, Interpretation—Social Studies. Untitled article discussing Russian territorial growth. Entire article was used.

For Unit 6. Interpretation—Natural Science. Article discussing forms of precipitation. Entire article was used, with some padding.

Note: All articles contained approximately 500 words each.

APPENDIX E

A SMALL SEGMENT OF A TYPICAL GSR RECORDING ON A STRIP CHART. GAIN SETTING AT 5.



REFERENCES

Books

1. Bacon, Francis, *The Complete Essays of Francis Bacon*, New York, Washington Square Press, Inc., 1963.

2. Hilgard, Ernest R., *Theories of Learning*, New York, Appleton-Century-Crofts, Inc., 1956.

3. Mowrer, O.H. *Learning Theory and the Symbolic Process*, New York, John Wiley and Sons, Inc., 1960.

4. Skinner, B.F., *Science and the Human Behavior*, New York, The Macmillan Company, 1953.

5. Wechsler, David, *The Measurement of Emotional Reactions; Researches on the Psychogalvanic Reflex*, New York, 1925.

6. Woodworth, R.S., and Schlosberg, Harold, *Experimental Psychology*, New York, Hold, Rinehart and Winston, 1962.

Articles

7. Balken, E.R., "Affective, Volitional, and Galvanic Factors in Learning," *Journal of Experimental Psychology*, VXI, 1933, 115-128.

8. Beamer, G.C. and Ledbetter, Elaine W., "The Relation between Teacher Attitudes and the Social Service Interest," *Journal of Educational Research*, 50, 1957, 655-666.

9. Bettinghaus, E.P., "The Operation of Congruity in an Oral Communication Situation," *Speech Monograph*, 28, 1961, 132-142.

10. Burch, Neil R., Childers, Harold E., and Edwards, Robert J., "Automatic GSR Analyzer," *Technical Document, Reports No. 63-74* USAF School of Aerospace Medicine, 1963.

11. Burdick, Harry A., and Burns, Alan J., "A Test of 'Strain Toward Symmetry,' Theories", *Journal of Abnormal and Social Psychology*, 57, Nov. 1958, 367-370.

12. Callis, R., "Change in Teacher-Pupil Attitudes Related to Training and Experience," Educational Psychology Measurement, 10, 1950, 718-727.

13. Cofer, Charles, N., " The Psychogalvanic Response as an Indicator of Emotional Reaction to Personality Test Items," American Psychologist, 3, 1948, 303.

14. Cooper, J.B. and Pollock, D., "The Identification of Prejudicial Attitudes by the Galvanic Skin Response, "Journal of Social Psychology, 50, 1959, 242-245.

15. Edelberg, R. and Burch, N.R., "Skin Resistance and Galvanic Skin Response, "Archives of General Psychology, VII, 1962, 163-169.

16. Esper, E.A. and Fairfax, V., "The Relation of Electrodermal Resistance to Performance in a Serial Learning Task, "Psychological Bulletin, 39, 1942.

17. Fere', C., "Nate sur les Modifications de la Resistance Electrique sous l' Influence des Excitations Sensorielles at des Emotions," E. R. de la Soc. de Biologie, XL, 1888, 217. (From Wechler, D., The Measurement of Emotional Reactions.)

18. Getzels, J.W. and Jackson, P.W., "The Teacher's Personality and Characteristics," Handbook of Research on Teaching, Chicago, Rand McNally and Co., 1963.

19. Gopalaswami, M.V., "A Note on the Correlation between the Psychogalvanic Reflex and 'Learning Effort', Indian Journal of Psychology, 1, 1926, 35-38.

20. Griggs, A.E., "A Validity Study of the Semantic Differential Technique," Journal of Clinical Psychology, 15, 1959, 179-181.

21. Gulliksen, H., "How to Make Meaning More Meaningful," Contemporary Psychology, III, 1958, 115-118.

22. Hovland, E.I. and Rosenber, M.J., "Attitude Organization and Change," Communication and Persuasion, Hovland, Janis and Kelley, New Haven, Yale University Press, 1953

23. Haggard, E. and Jones, H.D., "The Comparative Discriminatory Value of Various Measures of GSR for Words of Different Affect Value," American Psychologist, 1947, 349.

REFERENCES (Continued)

24. Iowa Tests of Educational Development, Chicago, Science Research Associates, Inc., 1962.

25. Jenking, J.J.; Russell, W.A.; and Suci, G.J., "An Atlas of Semantic Profiles for 360 Words," American Journal of Psychology, 71, 1958, 688-699.

26. _____, "A Table of Distances for the Semantic Atlas," American Journal of Psychology, 72, 1959, 623-625.

27. Kaplan, H.R.; Burch, N.R.; Bloom, S.W., and Edelbery, R., "Affective Orientation and Physiological Activity (GSR) in Small Peer Groups," Psychosomatic Medicine, 1963, 242-252.

28. Kleinsmith, L.J. and Kaplan, S., "Paired Associate Learning as a function of Arousal and Interpolated Interval," Journal of Experimental Psychology, 65, 2, 1963, 190-193.

29. Kuppers, W., "Higher Mental Processes and Galvanic Skin Response," Z. Exp. Angewand Psychol., 2, 1954, 291-32).

30. Lacey, Oliver L., "An Analysis of the Appropriate Unit for Use in the Measurement of Level of Galvanic Skin Resistance," Journal of Experimental Psychology, 37, 1947, 449-457.

31. _____, and Siegel, Paul S., "An Analysis of the Unit of Measurement of the Galvanic Skin Response," Journal of Experimental Psychology, 39, 1949, 122-127.

32. Leeds, C.H., "A Scale for Measuring Teacher-Pupil Attitudes and Teacher Pupil Rapport" Psychological Monograph, 64, 6, 1940.

33. Livonian,E., "Measure and Analysis of Physiological Response to Film," Title VII, Project No. 458, National Defense Education Act of 1958, Gr. No. 704094.

34. Lykken, D.T., "The Validity of the Guilty Knowledge Technique; the Effects of Faking," Journal of Applied Psychology, 44, 4, 258-262.

35. McKinney, F., "Certain Emotional Factors in Learning and Efficiency," Journal of General Psychology, 1933, 101-116.

36. McLeary, R.A., "The Nature of the Galvanic Skin Response," Psychological Bulletin, 47, 1950, 97-117.

37. Newcomb, T.M., "An Approach to the Study of Communicative Acts," Psychological Review, 60, 6, 1953, 393-404.

38. Niimi, Y., and Hashimoto, H., "Experimental Studies on Galvanic Skin Response Centered on the Unit of Measurement and Diminution Effect," Japanese Journal of Psychology, 24, 1953, 29-39.

39. Noble, E.C., "Meaningfulness and Familiarity," Verbal Behavior and Learning, Charles N. Cofer and Barbara S. Musgrave (Eds), New York, McGraw-Hill Co., Inc., 1963.

40. Rachman, S., "Reliability of Galvanic Skin Response Measures," Psychological Report, 6, 1960.

41. Ringness, T.A., "GSR During Learning Activities of Children of Low Average, and High Intelligence," Child Development, 33, 4, 1962, 879-889.

42. Silverman, A.J., Cohen, S.I., and Shmavonian, B.M., "Investigation of Psychophysiologic Relationships with Skin Resistance Measures," Journal of Psychosomatic Research, 4, 1959, 68-87.

43. Tarchanoff, J. "Euber die Balv. Erscheinungen an der Haut des Menschen bei Reizung der Sinnosorgane und bei Verscheidenen Forman der Psychischen Tatigkeit," Pflugers Archiv., VLIV, 1890, 46, 56, (From David Wechsler, The Measurement of Emotional Reactions.)

44. Veraguth, O. and Brunschweiler, "Recherches sur le Phenomene Psychogalvanique dans Quelques cas de Troubles Sensitifs par Blessures Cerebrales de querre," Revue Neuro., XXIV, 1918, 151-162. (From David Wechsler, The Measurement of Emotional Reactions.)

45. Weinrich, U. "Travels Through Semantic Space," Word, 14, 1958, 346-366.

46. Osgood, C.E., Suci, G.J., and Tannenbaum, P.H., The Measurement of Meaning, Urban: University of Illinois Press, 1957.